MULTIPLE THESAURI IN ONLINE LIBRARY BIBLIOGRAPHIC SYSTEMS

ECH SER
95
I22
087

ging Distribution Service, Library of Congress

Boston Public Library
Copley Square

MULTIPLE THESAURI IN ONLINE LIBRARY BIBLIOGRAPHIC SYSTEMS

A Report Prepared for
Library of Congress
Processing Services

by

Carol A. Mandel

Cataloging Distribution Service
Library of Congress, Washington, D.C., 1987

Z695
.M22
1987

Library of Congress Cataloging-in-Publication Data

Mandel, Carol A.
Multiple thesauri in online library bibliographic systems.

Bibliography: p.
1. Thesauri. 2. Subject cataloging—Data processing. 3. Subject headings. 4. Machine-readable bibliographic data. 5. Bibliographical services—Data processing. 6. On-line bibliographic searching. 7. Catalogs, On-line. I. Library of Congress. Processing Services. II. Title.
Z695.M22 1987 025.4'9'0285 87-3997

For sale from: Customer Services Section
Cataloging Distribution Service
Library of Congress
Washington, D.C. 20541 U.S.A.
(202) 287-6100

Table of Contents

	Acknowledgements	ii
	Executive Summary	iii
1.0	**Overviews: Multiple Thesauri and Computer Support for Controlled Vocabularies**	1
1.1	Introduction: A Multiplicity of Thesauri	1
1.2	Multiple Thesauri in Library Catalogs	1
1.3	Computer support for controlled vocabularies	2
2.0	**Thesaurus Management**	6
2.1	Computer support for thesaurus management	6
2.2	Computer support for relating multiple thesauri	8
3.0	**Subject Authority Control**	11
3.1	Background	11
3.2	ORION	11
3.3	WLN Computer System	16
3.4	DOBIS as Implemented by the National Library of Canada	31
3.5	UTLAS	43
3.6	Geac Bibliographic Processing System	52
3.7	NOTIS (Northwestern Online Total Integrated System)	55
3.8	Carlyle Systems TOMUS	66
4.0	**Subject Searching**	69
4.1	Online catalog search features	69
4.2	Approaches to Retrieval of Multiple Subject Vocabularies	70
5.0	**Support for Multiple Thesauri at the Library of Congress**	80
5.1	Current Situation	80
5.2	Computer Support for Thesaurus Management	87
5.3	Online Subject Authority Control	89
5.4	Online Subject Searching	90
	Bibliography	93

Acknowledgements

This report addresses a topic not yet documented in the literature. Thus the chief sources of information for many of the following sections are interviews—interviews with thoughtful, dedicated professionals who are shaping tomorrow's systems for subject access. Many knowledgeable individuals contributed to this paper. For taking the (considerable) time to describe their systems to me, I am deeply grateful to: Pat Moholt, Toni Peterson, and Cathy Whitehead of the Art and Architecture Thesaurus; David Bearman, Richard Szary, and Stephen Toney of the Smithsonian Institution; Diane Bisom, ORION; Susan Jurist, RLIN; Scott Valentine, NLC/DOBIS; Andrew Herkovic and Joanna Rood of UTLAS; Michael Ridgeway, Geac; JoAnne Kingsbury, UCSD (BLIS). I also wish to thank the many patient LC staff members who provided information to me, including: Linda Arret, John Graves, John James, Shirley Loo, Jane Marton, Mary Lou Miller, Barbara Orbach, Harriet Ostroff, Elisabeth Parker, Mary Kay Pietris, Ann Ross, David Smith, Alan Virta, and Helena Zinkham. I hope my report will prove useful to them. I am also thankful for the help of Julia Blixrud who interviewed many vendors of automated authority control systems to assess their ability to support multiple subject authority files and who prepared the descriptions of NOTIS and TOMUS found in Chapter Three. Finally, a grateful acknowledgement is due Cathy Marzec for her able assistance in typing the report.

Executive Summary

1.0 Overviews: Multiple Thesauri and Computer Support for Controlled Vocabularies

While most North American libraries have standardized their controlled subject vocabulary by using the Library of Congress Subject Headings (LCSH), separate catalogs using specialized thesauri are also common for certain subjects (e.g., medicine, art), materials (e.g., graphic materials), and audiences (e.g., children). As these special catalogs are automated along with main library card catalogs, libraries are faced with the challenge of providing computer support for multiple subject lists.

Computer support for controlled vocabularies entails support for three functional areas: 1) thesaurus management, 2) subject authority control, and 3) subject searching. The three areas are closely related and can be integrated in a system where the thesaurus is maintained as an authority file and is indexed and linked to bibliographic records for searching. ISIS serves as an example of such an integrated system in which the subject thesaurus plays a central role. In general however, most systems focus on only one of the three functional areas, and the current state-of-the-art in computer support varies greatly among the three.

2.0 Thesaurus Management

In general, computer systems for the creation, maintenance, and production of controlled vocabularies fall into two categories: 1) those designed for the production of a published thesaurus, and 2) those designed for subject authority control. A thesaurus management system for LC should include the best features of both kinds of applications and should provide a MARC format output of thesaurus term files. Desirable features include: online input and editing, support of coding and content designation, creation and subsequent authorization of provisional terms, records for non-postable terms, links from term records to references in other term records, links between term records, manipulation and displays of hierarchies, flexible alphabetical displays, flexible output specification, creation of microthesauri, creation of sets of related terms.

Information scientists have begun to explore the possibilities for creating automated systems that help to relate separately created thesauri to each other. Techniques for relating vocabularies include: mapping; creation of an intermediate lexicon, or switching language; integrating vocabularies in a master list; creating microthesauri; and creating a macrovocabulary such as UNISIST's Broad System of Ordering. Vocabulary integration can also be done as part of the editorial process of maintaining each of the vocabularies in question. This editorial work of integration could be facilitated by computer support if the vocabularies in question

were maintained on a system that provided such features as: checking new candidate terms against postable and non-postable terms in all lists; specification of term compatibility among lists; specification of cross-list relationships; merged and separate term displays for one or more lists.

3.0 Subject Authority Control

The state-of-the-art in subject authority control is well developed, with many systems operational that support online validation, maintenance, and syndetic structure for subject headings in a file of bibliographic records. A survey and telephone interviews identified at least seven systems that currently can provide authority control for more than one logically separate subject vocabulary: ORION, WLN, NLC/DOBIS, UTLAS, Geac BPS, NOTIS, and Carlyle.

The seven systems maintain subject authority records that include identification of subject source list, and validate subject headings against selected vocabularies. The systems vary considerably in their design and functionality. Validation ranges from a match of headings in order to capture cross-references (e.g., Carlyle) to complex linking mechanisms that replace headings in bibliographic records with authority record i.d. numbers (e.g., WLN, UTLAS). In some systems, validation/linking can occur on parts of headings, allowing for control of main terms separate from subdivisions. A few systems support the construction of links between authority records (e.g., through 5xx fields), and some can support these links between records from different subject source lists. For example, NLC/DOBIS and UTLAS create a special translation link between LCSH terms and their French counter-parts in Repertoire de vedettes-matiere.

In some systems, such as NOTIS, separate indexes are created for terms from different source lists, while in others terms from all vocabularies are retrieved together in a subject search, usually with a code identifying the source of non-LCSH terms. In other systems, such as Geac's BPS, profiles and file definitions can determine how subjects from different source lists will be indexed and displayed.

4.0 Subject Searching

The extent to which multiple controlled vocabularies can be used effectively in a particular system will be governed in part by the basic searching features of that system. Relevant features include: 1) index construction (e.g., component words vs. exact phrases), 2) displays (i.e., how easy is it to interpret results), 3) subject indexing browse (the possibility of selecting a term from a list), and 4) use of a subject authority file (the possibility of viewing cross-references and related terms).

There are four basic approaches to providing access to databases indexed by different vocabularies. These can be characterized as follows:

1. Segregated files. In this approach, collections using different subject thesauri are searched separately. The advantage of this approach, which works best when the different vocabularies are applied to distinctly different materials, is that it enables searchers to take full advantage of the specially developed features of each vocabulary. A disadvantage is the need for multiple subject searches when users require material from more than one collection.

2. Mixed vocabularies. In this most commonly used approach, terms from all vocabularies are retrieved together in subject searches. The two major problems caused by this approach are: 1) obvious vocabulary clashes (e.g., the same term is postable in one vocabulary and non-postable in another), and 2) degradation of access to specialized collections. The seriousness of these problems depends upon the subject searching design features of the online catalog and the particular mix of collections and vocabularies included.

3. Integrated vocabularies. Techniques used to relate different thesauri (described in Chapter Two) could be used to develop syndetic structures that would aid retrieval in those online catalogs that employ an authority file in subject searching. This approach would require both machine-assistance and human editorial work to integrate the vocabularies in question.

4. Front-end navigation. Features being developed in "intelligent" interfaces to online catalogs could be designed specifically to aid in searching from multiple vocabularies. These features range from sequences of algorithms for term matching to help screens designed to suggest vocabulary alternatives.

5.0 Support for Multiple Thesauri at the Library of Congress

In addition to maintaining and applying the 145,000-term LCSH, the Library of Congress has developed several specialized subject vocabularies tailored to the needs of specific applications (e.g., a printed bibliography) or special collections. These are: the Legislative Indexing Vocabulary (LIV), the LC Thesaurus of Graphic Materials (TGM), Subject Headings for Children's Literature, terms used in the index of the National Union Catalog of Manuscript Collections (NUCMC), and terms used in the index to the Handbook of Latin American Studies (HLAS). LIV and TGM are maintained as independent thesauri, meeting ANSI Z39.19 standards. The Children's List and NUCMC terms are LCSH-based, but include variations in terminology,

subdivision practice, and rules for application. The HLAS terms are maintained in the published index.

Both LEXICO and LCSH Online are used for thesaurus management at LC. Enhancements to one or both systems would facilitate management of the Library's thesauri. LEXICO has proven useful for LIV and TGM, and could be used for the HLAS list as well. However, it lacks a MARC Authorities Format output, so that thesauri created on LEXICO cannot be loaded into other LC files or used as subject authority files in other LC systems. Other limitations in LEXICO keep it from being fully functional for LIV, requiring duplicative input of LIV into two systems. LCSH Online, on the other hand, lacks lexicographical features that would ease and enhance the work of LCSH editors. Additional enhancement to LCSH Online would be desirable to enable cost-effective maintenance of the Children's and NUCMC exceptions in MUMS. Finally, if LC were to undertake some degree of integration among a number of its vocabularies, it would be desirable to maintain the linked thesauri on the same system.

In looking at its needs for subject authority control, a key decision for LC is whether to embark on significant redesign or replacement for MUMS and SCORPIO. If redesign/replacement is considered, the systems described in Chapter Three demonstrate that authority control of terms from multiple subject lists need not pose a problem. In fact, many of the functions and file structures that support other desirable subject authority control features--e.g., subdivision validation, global change on subfields, links between authority records--are those which help to support the creation of multiple subject authority files. UTLAS and Geac BPS are two powerful systems that provide good models for LC's subject authority environment.

No operational system has yet made the searching of terms from multiple thesauri in the same file easy for the untrained end-user. In determining how it will address this retrieval problem in its own catalogs, LC will need to determine which of the four approaches outlined in Chapter Four it wishes to pursue. One possibility would be to develop a strategy comprising both short term and long-term plans. For example, a short term "solution" might be to segregate the different vocabularies, or to continue to tolerate the present mix. At the same time, work could begin on a longer term plan, such as integrating the vocabularies (an ongoing editorial effort) or designing a retrieval "front-end" that guided users through multiple vocabulary searching. This latter approach would entail a significant research and development effort that could benefit other libraries as well as LC.

Chapter One
Overviews: Multiple Thesauri and Computer Support for Controlled Vocabularies

1.1 Introduction: A Multiplicity of Thesauri

The controlled vocabularies used for subject indexing are as many and as varied as the materials they index and the audiences they serve. Each time a specialized catalog, index, or information service is developed, the selection or creation of a suitable thesaurus is a critical decision. Recently, Stam (14) studied the factors that influenced this decision in the design of some 20 art historical information systems. She concluded that, while the purpose of the information system was certainly the primary determinant in selecting a thesaurus, a number of environmental factors also contributed significantly to the decision process. She characterizes these as: source of funding; nationalism and national language; history; tradition: habit, preference, and convention; convenience; institutional structure; and accident.

Noticeably absent from Stam's list is the desire or motivation to standardize. Art museums have had little impetus to try to share common vocabularies. The same may be said of abstracting and indexing (A&I) services. Most A&I tools developed independently, each within the context of its own materials, operations, and clientele. Fifteen years ago, when each tool was printed, published, and perused separately, the specialized vocabulary of each did not appear to pose a problem for its users. The situation is significantly different in today's world of online access. By 1985, more than 2,500 databases were offering bibliographic citations through some 360 online services. (17) When a service such as DIALOG offers 200 diverse databases through a single search interface, searchers begin to demand assistance in moving among the multitude of thesauri available online. A variety of techniques to address this problem have been tried, such as an integrated index or vocabulary switching, but no satisfactory solution is fully operational. Specialized thesauri, each designed to provide the best possible subject access within its own application, begin to pose retrieval problems when multiple files must be searched.

1.2 Multiple Thesauri in Library Catalogs

Unlike abstracting and indexing services, North American libraries have traditionally taken a standardized approach to the provision of subject indexing for their clientele. This standardization is motivated by two factors that, for the most part, do not operate in the worlds of art museums or A&I services. The first is the availability of standardized source copy for the bibliographic records used in library catalogs, source copy containing controlled vocabulary terms. The second

factor is the broad subject coverage and general user focus of most library catalogs. One result of these conditions is that thousands of North American library catalogs share a common thesaurus, the Library of Congress Subject Headings (LCSH).

At the same time, libraries have also supported a variety of specialized catalogs, typically in branch or departmental collections. Separate catalogs for medicine, art, music, law, children's collections, and visual materials are common. Many of these catalogs employ specialized thesauri, some of which have become national standards within their specialized fields. In university libraries, for example, the National Library of Medicine's Medical Subject Headings (MeSH) is a common "second" thesaurus. These specialized collections are beginning to join the ranks of other materials given standard machine-readable cataloging. For example, the last few years have seen the development of MARC formats for a variety of non-book material. As new formats and special collections are cataloged, the thesauri used to index them have begun to appear in MARC format databases. Peterson (12) notes that between the publication of MARC Formats Updates 10 and 11, the allowable codes indicating subject term source lists had grown from four to nineteen lists.

As libraries bring specialized catalogs online, the searching of library catalogs begins to replicate, on a more modest scale, the searching of multiple databases. While most libraries have far fewer than 200 vocabularies available in their online catalogs, the problem is more potent for libraries than for other database providers. This is because online catalogs are intended for direct use by library patrons. Further, libraries are concerned with the continuing maintenance of subject terms and references within their online catalogs, an expensive effort further multiplied by the number of controlled vocabularies contained in the database. A relatively few library bibliographic systems have been designed to support a library in maintaining more than one controlled vocabulary in its catalog. (These are described in Chapter Three.) None has yet been employed specifically to assist patrons in retrieval from multiple thesauri.

1.3 Computer support for controlled vocabularies

The maintenance and application of controlled vocabularies can be made more effective with the application of appropriate automated support. In library catalogs and cataloging, this encompasses three broad functional areas.

1. Thesaurus management. This functional area includes the creation, maintenance, and production of a list of controlled vocabulary terms. While many libraries do not maintain their own local subject lists, support for this function is particularly important at the Library of Congress where LCSH and several other vocabularies originate.

2. Subject authority control. This functional area includes the verification and maintenance of subject headings and references in a file of bibliographic records.

3. Subject searching. This area includes functions that support the retrieval of bibliographic records through inquiries on controlled subject terms. Subject searching in a bibliographic file may or may not include access to a subject authority file.

The work of subject analysis could also be considered as a fourth functional area. However, at the Library of Congress and most libraries, this work is largely intellectual. To the extent that computer support can be provided to aid in subject analysis (e.g., displays that facilitate the selection of appropriate terms), functions desirable in the other three areas are also those functions that meet the needs of subject catalogers. Given libraries' large investment in subject analysis work, selected features in each area that are considered of highest priority by subject catalogers merit particular attention.

The three functional areas are closely related, and automated systems designed to support any one of them will necessarily include some features that support another. For example, most subject authority control systems support the creation and updating of subject authority records, an aspect of thesaurus maintenance. Subject searching must also be supported in some manner (not necessarily a user-friendly manner) in order to access terms in a thesaurus or subject authority file. Despite these interrelationships, most controlled vocabulary support systems have initially been developed with an emphasis on only one or another of the three areas. Sophisticated thesaurus management systems tend to focus on the subject list itself, and not the creation of a subject file that can also be linked to bibliographic records. Currently, the most sophisticated authority control systems are not those available for end-user searching, and operational online catalogs to-date have made little use of a subject thesaurus as a retrieval aid. However, a trend can be perceived toward a more integrated approach to subject support. LCSH Online (MUMS Subject Release 1.0) forms the core of a thesaurus maintenance system that supports the creation of a subject authority file. In the next version of its system, the Art and Architecture Thesaurus plans a MARC Authorities Format output that could easily serve as a subject authority file loaded into an appropriate bibliographic system. Online catalog designers are exploring ways to enhance subject searching through the use of thesaurus displays, and user-friendly "front ends" are being developed for cataloging systems that have powerful subject authority control (for example, the BLIS EZ Access interface to WLN). The 1981 ASO Task Definition for LCSH and LIV Online took a similar, integrated view of subject support, but subsequent development at LC has veered temporarily from this course.

ISIS (Integrated Set of Information Systems), along with its minicomputer version MINISIS, is a system that provides a conceptually interesting example of a bibliographic system developed with an emphasis on support of a controlled vocabulary. Designed for use in developing countries by the International Development Research Centre, ISIS was created to provide multilingual access to bibliographic records. The system supports the creation and maintenance of a thesaurus/subject authority file in which the record for each term can contain links to: non-postable terms, broader terms, narrower terms, related terms, and translations of the term in up to nine languages. Edit checks can be done to confirm that necessary reciprocal main terms exist in the thesaurus for all references and translations. As inverted files (indexes) for bibliographic records are built, index terms are validated against the thesaurus, and bibliographic records are linked to thesaurus records (i.e., subject authority control is performed). In searching, the thesaurus serves as the inverted file for the bibliographic database so that queries can be matched against translated terms and references. Searching is designed to help the users select terms; a query may request broader, narrower, or related terms. For example, a search for "School" and broader terms might display the following terms and postings:

```
Q7  BT  School
    School          p = 3
    Ecole           p = 3
    Escuela         p = 1
Educational institution          p = 79
Etablissment d'enseignement      p =  3
Establecimiento de ensenzana     p =  2
```

ISIS also supports the creation of tables of related terms linked to a broad, "ANY," term (similar in concept to "bucket terms" in LIV). For example, the terms "China," "Korea," and "Japan" might all be linked to the ANY term "Far East." Queries requesting an ANY term will retrieve the list of related terms.

ISIS does not support the creation and loading of MARC format-compatible bibliographic and authority records, thus limiting its usefulness in U.S. library applications. However, in conceptualizing automated support for a controlled vocabulary, ISIS illustrates a design in which the controlled vocabulary is at the core of the system. The three functional areas of thesaurus management, subject authority control, and subject searching are all equally important in the effective application of the system.

In examining its own needs for computer support of locally-created thesauri, the Library of Congress will need to consider support in each of the three broad functional areas. While the goal of a single integrated system such as ISIS is probably not realistic (or even desirable) in LC's complex environment, support for vocabulary control should be integrated to the extent that work is not duplicated (e.g., each thesaurus record

should not need to be keyed more than once) and files developed in one unit are easily available for online viewing in other areas where subject work is done. Additional questions of integration must be considered in developing an automation strategy for LC's multi-thesaurus environment: Can separate parallel thesaurus management systems be employed for each thesaurus or is some degree of integration of the separate thesauri necessary? Should subject searching on the different thesauri be integrated? Here again, LC's complex operational environment would indicate that a single fully integrated system may not prove to be an attainable goal.

Outside the Library, the current state-of-the-art in system support varies greatly among the three functional areas: thesaurus management systems tend to be specialized or limited applications; subject authority control is well developed in a number of bibliographic systems; end-user subject searching is in an early but rapidly developing state, with no system yet operational that effectively aids the user in searching terms from multiple controlled vocabularies. For this reason, the following chapters address each of the functional areas separately. The different organizations and varying kinds of information presented in each chapter reflect the differences in state-of-the-art automated support for each.

Chapter Two
Thesaurus Management

2.1 Computer support for thesaurus management

In general, computer support for the creation, maintenance, and production of controlled vocabularies has been developed in two kinds of operational environments. In the first environment, the creation of the controlled vocabulary is viewed as an end in itself; typically the end product is a printed thesaurus. In the second, the vocabulary is maintained as a subject authority file linked to bibliographic records.

Applications focused on the thesaurus itself have engendered the development of a wide variety of specialized software packages for thesaurus management in organizations all over the world. These applications include a host of functions that aid the lexicographer in selecting terms, building hierarchical and other relationships, and fine-tuning the list. Some thesaurus management systems are designed for the development of multilingual thesauri; others are tailored to the handling of facets. Research has even begun on the potential of artificial intelligence in thesaurus software, in this case the use of a high-level "object-oriented" programming language in thesaurus management systems. (4)

Authority file applications focus on coding terms and references for interaction and searching in a machine file--often a file compatible with MARC bibliographic and authority records. A number of subject authority systems support the creation and editing of more than one controlled vocabulary, along with the ability to create references across subject lists. These systems are described in detail in Chapter Three.

Since the Library of Congress is responsible both for the production of published thesauri and for the use of the same controlled vocabularies in a bibliographic system, a thesaurus maintenance system for LC should contain appropriate features from both kinds of applications. These features, culled from a review of various systems, can be summarized as follows:

1. Online input and edit. These features include convenient creation and editing of records and references (screen prompts, word processing), checks for duplication as new terms are entered (against both postable and non-postable terms), error checks for appropriate coding and content designation.

2. Support of coding and content designation for authorized term records. For maximum usefulness, each record for a main thesaurus term should contain all data recommended in the ANSI Z39.19-1980 Standard for Thesaurus Construction and the U.S. MARC Format for Authorities. Additionally, thesaurus records may contain class numbers for the thesaurus term classification.

3. Provisional headings. In the editorial process of thesaurus maintenance, it is often necessary to designate new terms as candidate or provisional terms and to be able to retrieve, display, and output these terms for editorial review and approval. New references may also require the same process.

4. Records for non-postable terms. Separate records are often useful for non-postable terms, particularly if these were formerly main terms. Such records can contain information on why the term was rejected or fell out of use.

5. Links from term records to references in other term records. These features include: a) automatic checking for the existance of a record for any term chosen as a BT, NT, or RT; b) automatic creation of reciprocal references (e.g., BT/NT relationships, RT relationships, translated terms) when the first half of the pair is posted; c) automatic maintenance of a term used as a reference in another record whenever the term is changed; d) ability to display a term along with all terms related to it.

6. Links between term records. Such links provide the ability to link a term to a record for a subdivision or geographic heading. (For a description of this feature in an operational system, see the section on UTLAS in Chapter Three.)

7. Manipulation and displays of hierarchies. These features include the ability to view, manipulate, and edit terms within hierarchies; to display different levels of hierarchy; to view simultaneous displays of the same term used in different hierarchies; and to output hierarchical listings.

8. Flexible alphabetical displays. These displays include a left match alphabetical browse of terms (with and without references), and KWIC and KWOC displays.

9. Flexible output specification. Since most thesauri are published or distributed in some print or micro form, the ability to specify print tapes for a variety of products is necessary. Also critical for bibliographic system use is a machine-readable output in the MARC Authorities communications format. It should be possible to output the entire file, updates only, and specified subsets of the file (see item 10).

10. Microthesauri. Many systems support the creation of specified subsets of the file, or microthesauri.

11. Bucket terms. Some systems support the creation of sets of related terms linked to a parent term (e.g., "bucket" terms in LIV, "ANY" terms in ISIS).

No existing system supports all of these features. This is largely because the most sophisticated thesaurus management systems do not support the MARC authorities format and cannot be linked to a standard library file of bibliographic records.

2.2 Computer support for relating multiple thesauri

The features just described are useful for managing one or more than one thesaurus when each thesaurus is maintained independently. Additional computer support is necessary when one attempts to achieve some degree of compatibility or integration in a multiple thesaurus environment. Such efforts have received considerable discussion in the recent literature of information science, largely in response to the difficulties encountered by online database searchers faced with a wide variety of specialized thesauri. Vocabulary compatibility is an issue of particular concern to those involved with international information exchange, since language adds yet another dimension to the complexities of searching across vocabularies. Lancaster and Smith provide an excellent review of the literature in a report prepared for UNESCO's General Information Program. (5) For the purposes of their summary, they group approaches to thesaurus integration into five categories:

1. Mapping. This approach entails the direct translation of terms in one vocabulary to corresponding terms in another. Mapping can be in one direction, or reciprocal between two vocabularies.

2. Intermediate lexicon/switching language. This approach entails the mapping of two or more vocabularies to an intermediate or neutral language, such as a classification or coding scheme.

3. Integrated vocabulary. This approach entails the construction of a master list of all terms and references in the targeted vocabularies. The master list differs from a map or an intermediate lexicon in that it may not try to relate all vocabulary terms, but only those that automatically match.

4. Microthesauri. This approach treats specialized thesauri as satellites of a larger, broader hierarchy. However, it exists more in theory than reality, since most specialized thesauri are created independent of a broader list.

5. Macrovocabularies. This approach entails the creation of a generic superstructure to which terms in specialized thesauri are linked. The best known application of a macrovocabulary is the UNISIST Broad System of Ordering (BSO), which is a classification scheme of some 4,000 terms with which other vocabularies can be coded.

Vocabulary integration is a fertile area for computer applications, and researchers report a wide variety of experimental systems. Machine support is particularly valuable for mapping terms and merging term lists, with algorithms developed for matching on references, component words, stems, boolean combinations, etc. Machine-assistance for integration is most useful when certain levels of compatibility exist in the maintenance of the lists, particularly in the use of common

standards for constructing the list and compatible record structures for encoding terms. In international efforts, MATER (Magnetic Tape Exchange for Terminological/Lexicographical Records) is being forwarded as an exchange format for thesaurus term records. (13) Developed by the German Institute for Standardization, MATER has been elaborated as ISO DP 6156. In the U.S., the MARC Format for Authorities clearly fills that bill. No machine system can eliminate the need for extensive intellectual effort in relating terms, particularly when vocabularies differ in their levels of specificity, use of hierarchical structure, and level of precoordination.

Most of the systems designed to aid vocabulary integration reflect an effort to map separately maintained vocabularies after they are created, i.e., for retrieval. In a library environment where a very limited number of vocabularies are created and maintained, it is also possible to imagine a limited effort at mapping or integration at the point of editing the controlled vocabularies. For example, the Art and Architecture Thesaurus is attempting compatibility with LCSH by coding all LCSH terms used in the AAT and providing USE references whenever an LCSH term is not selected. In a situation where such compatibility might be maintained, it would be desirable to employ thesaurus management software that supported the development of relationships across lists. A number of the subject authority systems described in Chapter Three have features that can support this kind of integrated approach to thesaurus maintenance. Desirable features would include the following:

1. Ability to view and maintain each thesaurus separately, permitting editorial consistency within each list.

2. Automatic checking of new candidate terms against postable and non-postable terms in all thesauri. Ideally, checking would include a variety of options for matching terms, e.g., against stems, component words, etc. so that new terms were thoroughly linked in all lists.

3. Ability to specify compatibility of terms (e.g., both LIV and AAT term records include codes for compatibility with LCSH), especially for terms which are postable in one list and non-postable in another.

4. Ability to copy a term record from one thesaurus to another without rekeying; alternatively, the ability to code a single term record for use in more than one thesaurus.

5. Ability to specify and establish cross-file relationships; usually this is an RT, but a specific "equivalent term" relationship might be desirable. (For example, translation relationships are specified in NLC/DOBIS and UTLAS.) Many of the linking features described earlier for a single thesaurus would be desirable in cross-file links, particularly the automatic maintenance of cross-references that are also main terms in other lists.

6. Ability to search and display terms from one or more lists as specified by the user. Desirable displays of terms would include: merged alphabetical and KWIC displays, with terms coded to include their source thesaurus; parallel hierarchical displays to show the same term in hierarchies from different lists.

Chapter Three
Subject Authority Control

3.1 Background

The state-of-the-art in subject authority control is well developed, with a considerable number of systems operational that support validation, maintenance, and syndetic structure for subject headings in a file of bibliographic records. This is somewhat surprising given the recent unavailability of a MARC Format file of LCSH authority records. However, a number of utilities, such as UTLAS and WLN, have created subject authority files to serve the needs of cataloging customers; several newly developed online systems, such as Geac BPS and Carlyle, have been designed in anticipation of customer demand for subject authority records in online catalogs.

In 1985, Taylor et al. published an updated list of networks and vendor-supplied systems that provide authority control. (15) As a first step in developing the descriptions that are presented in this chapter, each of the providers listed by Taylor was contacted in order to determine whether its system could support online, interactive authority control for more than one logical subject authority file. In addition, representatives of two locally-developed systems were contacted (the National Library of Canada's DOBIS implementation and the University of California, Los Angeles Library's ORION) as these local systems were known to maintain multiple subject authority files.

The survey and telephone interviews uncovered seven online systems that currently provide authority support for multiple controlled vocabularies: ORION, WLN, NLC's DOBIS, UTLAS, Geac BPS, NOTIS, and Carlyle. In-depth interviews with system representatives revealed a wealth of sophisticated subject authority control features operational or in the final stages of testing and refinement. The systems are interesting not only for their support of multiple thesauri, but for their power to maintain control over headings and subdivisions and for their support of linking relationships among terms. Many of the systems provide functions that ease thesaurus maintenance as well as headings control. While most of the systems have the potential to enhance end-user retrieval via the authority file, no installation has yet done the intellectual work necessary to make subject searching through multiple authority files accessible to the library patron.

Seven of the systems are described in this chapter.

3.2 ORION

3.2.1 Background

ORION is the online integrated library system developed and operated at the University of California, Los Angeles. ORION

supports a variety of processing functions (ordering, serials control, etc.) and an online catalog, each operating from the same MARC Format bibliographic data base. Subject authority control functions are currently under development, with core subject authority features already implemented. ORION operates at a single installation, the Office of Academic Computing at UCLA. However, libraries other than UCLA are supported by the software; each ORION user is installed as a separate database and linked to the UCLA computer center by telecommunications lines. For example, the Library at the J. Paul Getty Center has loaded its database into ORION and uses the system for technical processing and an online catalog.

3.2.2 Description of the Subject Authority File

ORION supports a MARC-format subject authority file, with some additional local fields added to authority records to control filing, record retention, and other local processing. The ORION subject authority file is built from headings copied from bibliographic records as they are loaded/created and from authority data keyed into the system online. Subject terms from various source lists are integrated into a single file. As subject headings are copied from bibliographic records, their source indicator value is carried into the subject authority file along with the heading and retained in the second indicator position (a slight variation from the MARC authorities format). To date, no ORION subject authority records have been constructed from headings coded with both a second indicator value of 7 and a source shown in subfield 2; if necessary, these would be distinguished by mapping a value other than "7" into the authority file. If the identical term is used in two different source lists, two separate subject authority records will be created, each coded appropriately. ORION planners expect to support subject authority control for headings from LCSH, MeSH, Annotated Card Program, and local lists.

3.2.3 Links and references

Bibliographic records are linked to the authority file by inserting the appropriate bibliographic record i.d. numbers into a specified field in the ORION authority record (the "1011" field). This link, which is made automatically as bibliographic records are loaded or created, is based on matches to an entire heading. For subject headings, this match includes all subdivisions (i.e., the entire string) and the second indicator. This link supports maintenance of headings on bibliographic records through a change in the subject authority file. ORION designers plan to develop chained commands that will ease maintenance of different subject headings that contain a common term, such as a main term that is used with different subdivisions.

Authority file records will contain 4xx (see from) and 5xx (see also from) references. When a heading is loaded that matches a 4xx reference, an error report will print; when a 4xx match is keyed online, an error message will display on the screen. Matches to 4xx fields include the indicator for subject source, so that headings from different thesauri are authorized separately.

No links are built between records in the authority file; for example the addition of a 5xx field in one subject authority record will not generate a 3xx in another. Maintenance of "see also" and other related term references is intellectual; records for related terms are not linked in the authority file.

3.2.4 Retrieval and display

Separate indexes are built for the authority and bibliographic files, and different commands are used to access each index. The "FIND SUBJECT" command leads to a keyword search of subject headings in the bibliographic file, with no distinction among terms from separate thesauri. (In fact, search terms of more than one word can retrieve a record in which one of the words appears in one subject heading and one appears on another.) Users are encouraged to do subject searching using the "BROWSE SUBJECT" command which performs a keyword search on headings and references in the subject authority file. Boolean operators can be used in the search, along with right truncation and internal truncation (e.g., "WOM?N" will retrieve both "woman" and "women"). Search terms are normalized for capitalization and punctuation. The BROWSE command retrieves an alphabetical list of subject headings and references, along with the number of linked bibliographic records, i.e., hits for each heading. A sample display is shown in Figure 0-1. Cross references appear in the form, for example, "Women physicians SEE Physicians, Women," and are interfiled alphabetically.

The subject authority file index is not partioned by subject source; terms from different thesauri are retrieved in the same alphabetical list. ORION designers plan to distinguish terms from different lists by a user-friendly code that will display next to each term or reference. For example, a search on "Clinical psychology" might result in:

hits	term	source
25	Clinical psychology	(LCSH)
	Clinical psychology SEE Psychology Clinical	(Medical)
12	Pyschology, Clinical	(Medical)
	Psychology, Clinical SEE Clinical Psychology	(LCSH)

Since MeSH headings are used in UCLA's Medical Library, catalogers hope to be able to create "search also" references between conflicting LCSH and MeSH terms. A MeSH/LCSH mapping project currently underway at the National Library of Medicine will help in this task. A help screen is available that describes the different thesauri used by UCLA libraries. At present, there are no plans to provide the means to limit subject searches by source but this could be developed if it proved necessary. Since ORION subject authority control is still under development, and since ORION is constantly being enhanced in response to UCLA Libraries' needs, additional features to support multiple subject thesauri may be added in the future.

BROWSE COMMAND

ORION will respond with a brief list of all of the subject headings that match that search request, in alphabetical order across all bibliographic files. This brief list display will number the subject headings (or name headings, etc., as the case may be) so that they can be used with the retrieve command. The number of online ORION records will appear next, followed by the heading. Cross references will also appear on this brief listing in the form, for example: See, Holy SEE Holy See. The see references will be interfiled alphabetically among the headings. The current search and the number of results will appear at the top of the list and the options for the user's next step will appear at the bottom of the screen. For example:

```
CURRENT SEARCH:  BROWSE SU CIVIL WAR
-SUBJECT HEADINGS CONTAINING ENTERED BROWSE TERM(S) - 220 RESULTS.
      <-> NUMBER OF ONLINE ORION RECORDS CONTAINED IN EACH GROUP.

R1      1  Alabama--Politics and government--Civil War, 1861-1865.
R2      1  Angola--History--Civil War, 1975- --Participation, Cuban.
R3      1  Arkansas--Politics and government--Civil War, 1861-1865.
R4      1  California--History--Civil War, 1861-1865.
R5      3  Cambodia--History--Civil War, 1970-1975.
R6      1  Cambodia--History--Civil War, 1970-1975--Addresses, essays,
             lectures.
R7      2  Charleston, S.C.--History--Civil War, 1861-1865--Sources.
R8      7  China--History--Civil War--1945-1949.
R9      1  China--History--Civil War--1945-1949--Drama.
R10     1  China--History--Civil War--1945-1949--Fiction.
R11     1  Civil War.
R12     1  Civil War--Case studies.

*OPTIONS: -TYPE R1 (OR R2, R3 ...) TO RETRIEVE THE RECORD(S) IN A GROUP.
          -PRESS THE ENTER OR RETURN KEY TO SEE MORE OF THE LIST.
          -BEGIN A NEW SEARCH (E.G. FIN NT ..., OR B SU ...)

ENTER NEXT COMMAND
```

The user may then use the command RETRIEVE and an index number; this will trigger a search on the external record numbers attached to the browse record, thus retrieving all of the records in ORION which have that subject heading.

3.3 WLN Computer System

3.3.1 Background

The WLN System is an integrated library system designed and owned by the Western Library Network. The core system features cataloging, catalog maintenance, and searching facilities, with linked authority control. WLN maintains and operates the software for its own network in the Pacific Northwest. In addition, WLN licenses the software to fourteen other organizations, including two university libraries in the U.S., the national libraries of Australia, New Zealand and Singapore, and the British Library. A former library software company, Biblio-Techniques, Inc., developed an enhanced version of the WLN software (the Biblio-Techniques Library Information System, BLIS) and installed the system at several sites in North America. Subject authority control features are essentially the same at all WLN installations. BLIS users have been provided with EZ Access, the capability to develop special screens and command flows that serve as a "user-friendly" interface for searching the system, enabling BLIS to be used as an online catalog.

3.3.2 Description of the Subject Authority File

The WLN System supports the loading and online creation of U.S. MARC Format bibliographic records. These records are linked to a file of authority records that contain most of the data elements used in the MARC Authorities Format. A list of the topical subject and reference authority data elements used in WLN is provided in figure W-1. The LCSH references "See", "See from" "See also", and "See also from" are included.

The WLN System does not support loading of or interface with MARC Format authority records. (A system reconfiguration that will permit MARC authorities compatibility is scheduled for completion in 1989.) Once an authority file is loaded into a WLN system, changes to authority records must be keyed online. New authority records are added to WLN only by online keying and automatic generation from headings on bibliographic records newly added to the system.

The WLN Authority File is partioned into three parts: Names, Series, and Subjects. The subjects file is further subdivided by "vocabulary type" into the following categories: LCSH, Annotated Card Program, MeSH, NAL, "Other," NLC English (i.e., Canadian Subject Headings), and NLC French (i.e., Repertoire de vedettes-matiere). These categories are derived from the thesaurus source coding in the 6xx second indicator field of MARC format bibliographic records. The "Other" category groups all subject terms coded with "7" in the second indicator and does not distinguish among "Other" lists. Each WLN authority record carries a code for "vocabulary type" (i.e., source list). If the same subject term is used in more than one source list, a separate record will be created with each coded for respective source.

3.3.3 Links and references

As bibliographic records are added to a WLN system, all headings on the record are matched against the authority file. Headings that match are linked to the appropriate authority record by replacing the bibliographic record heading with the unique i.d. number (ISN, or Internal Sequence Number-Vocabulary) of the authority record. For headings new to the system, an authority record is created and the link established. Matches are based on the complete heading (e.g., full subject strings). Subject headings are matched both by term and source; a bibliographic record containing subjects from multiple sources can be linked to subject authority records from various sources, provided each subject heading carries the appropriate second indicator value.

The link of headings on bibliographic records to authority records supports headings maintenance through the authority file. When an authority record is changed, linked bibliographic records are updated automatically. Since the link is for complete strings, a single edit will not globally change every appearance of a term; for example, a change on the heading "Canada-History" will not affect the headings "Canada-History-War of 1812" or "Education-Canada-History." However, a search and display option specially designed to aid database maintenance can provide a listing of all the headings in which a particular term is used, so that each subject heading will be identified for correction.

Within the authority file, a limited set of linked reciprocal relationships can be defined. These are: see/used for, see also/refer from, and former name/later name. Reciprocity is further limited by data element type, e.g., a corporate name series must refer to/from another corporate name series. Each authority record containing such a reference must be matched by a reciprocal record; a reciprocal record is generated even for non-postable (see reference) terms. When references are added, the system generates reciprocal records and/or references automatically. Figure W-2 describes an example. The link is built by replacing the "tracing" for the related term in the authority record with the ISNV (authority record i.d. number) for that term. Thus, headings changes are automatically changed on reciprocal authority records. Reciprocal relationships for topical subjects must include specification of both the subject terms and their source list. Reciprocal links can be defined within and across the vocabulary type subdivisions of the subject authority file. For example, the following authority record would generate a "refer from" reciprocal for the MeSH heading "Clinical psychology" and an LCSH "see" reference record for Clinical psychology:

(Topical subject LC)	SUT-L	Clinical psychology
(Topical subject see also MeSH)	SAT-M	Psychology, Clinical
(Topical subject see reference LCSH)	SET-L	Psychology, Clinical

3.3.4 Retrieval and display

Although bibliographic and authority files share an index (known as the Key File), they cannot be searched together. The "Find" command will retrieve bibliographic records directly; the "Term" and "Browse" commands search the authority files. If a subject search is performed using the "Find" command, no cross references or other subject information, such as identification of the subject source list, will be displayed. The retrieval result will be a list of bibliographic records as shown in Figure W-3. The sample result includes, without distinction, records in which the search term "Psychology" was used as a main term and subdivision, and as both an LC subject and a MeSH heading.

Much more sophisticated subject searching is supported by searches of the authority file, although the resultant WLN displays are not end-user-friendly. (BLIS users may modify these displays by use of EZ Access. Some intend to develop displays, help screens, and references that will ease searching in multiple subject lists, but have not yet done so.) A "Term" search will retrieve a listing of all headings and references in which the term appears as a complete subfield; subjects from various source lists are included in the result, but sorted separately. Figure W-4 displays the result of a "Term" search on "Psychology;" Figure W-5 shows the results of the same search on "Psychology, Clinical." Headings coded "M" are MeSH; headings without a code are from LCSH. An asterisk (*) to the left of a heading indicates that it is a cross reference; a plus sign (+) indicates that there are references associated with the heading. The searcher may select a term (including references) for a follow-on bibliographic file search.

The "Browse" search will display an alphabetical listing of subject authority headings. The alphabetical sort of subjects includes separate sequencing for each vocabulary type, so subjects from different source lists are viewed separately. For example, all LCSH, A-Z are listed first, then Childrens Headings, A-Z, followed by MeSH, etc. The searcher can indicate which subject list is to be browsed by qualifying the command; an unspecified "Browse" command will lead to the LCSH sequence. Figures W-6 and W-7 display the results achieved by browsing the term "Psychology" in both the LCSH and MeSH sequences.

Several display options are available for authority records, as shown in Figure W-8. The complete display shows all related headings. Two commands have been designed that allow subject searchers to view related headings without displaying complete authority records. A searcher may select a subject heading from a search result and issue the command "EXPAND." This would result in a listing of the selected heading along with the terms used in its "see" and "see also" references, as shown in Figure W-9. Issuing the command "EXPANDALL" would retrieve all references, including "see from" and "see also from" headings, as

shown in Figure W-9. Unfortunately, the resultant display is arranged alphabetically without indication of the relationships among terms.

Component word search is not supported for subject searching except indirectly through the maintenance file described earlier (the key file). However, the first term of any subject subfield may be searched and right truncation can be applied. Exact term subject searching is also supported, with non-"a" subfields entered in any order following the "a" subfield. Subject searching on WLN is very powerful, but BLIS users implementing an online catalog have considerable work to do to help users take advantage of subject authority file searching. Several BLIS libraries plan to use both LCSH and MeSH headings in their online catalogs; at least one plans the use of some local terms, while a Canadian library intends to support both French and English language terms.

SUT		Topical Subject
		Not defined
	L	LC adult
	C	LC children's
	M	MeSH
	A	National Agricultural Library
	V	Other
	E	NLC - English
	F	NLC - French
	a	Topical subject
	b	Name following place
	x	General subject subdivision
	y	Chronological subject subdiv.
	z	Geographic subject subdivision

Cross References: (indicators, subfield codes same as Authorized Heading)

Former Name (FN) Cross References:

FNB	Corporate Name Series Former Name Cross Reference
FNC	Corporate Name Author Former Name Cross Reference
FNL	Conference/Meeting Name Series Former Name Cross Reference
FNM	Conference/Meeting Name Author Former Name Cross Reference

Later Name (LN) Cross References:

LNB	Corporate Name Series Later Name Cross Reference
LNC	Corporate Name Author Later Name Cross Reference
LNL	Conference/Meeting Name Series Later Name Cross Reference
LNM	Conference/Meeting Name Author Later Name Cross Reference

"See" Cross References:

SEB	Corporate Name Series See Cross Reference
SEC	Corporate Name Author See Cross Reference
SEG	Geographic Subject See Cross Reference
SEL	Conference/Meeting Name Series See Cross Reference
SEM	Conference/Meeting Name Author See Cross Reference
SEO	Personal Name Series See Cross Reference
SEP	Personal Name Author See Cross Reference
SES	Title Series See Cross Reference
SET	Topical Subject See Cross Reference
SEU	Uniform Title Heading See Cross Reference

"See Also" (SA) Cross References:

SAB	Corporate Name Series See Also Cross Reference
SAC	Corporate Name Author See Also Cross Reference
SAG	Geographic Subject See Also Cross Reference
SAL	Conference/Meeting Name Series See Also Cross Reference
SAM	Conference/Meeting Name Author See Also Cross Reference
SAO	Personal Name Series See Also Cross Reference
SAP	Personal Name Author See Also Cross Reference
SAS	Series Title See Also Cross Reference
SAT	Topical Subject See Also Cross Reference
SAU	Uniform Title See Also Cross Reference

Tracings: (indicators, subfield codes same as Authorized Heading)

"Used For" (UF) Tracings:

UFB	Corporate Name Series Used For Tracing
UFC	Corporate Name Author Used For Tracing
UFG	Geographic Name Used For Tracing
UFL	Conference/Meeting Name Series Used For Tracing
UFM	Conference/Meeting Name Author Used For Tracing
UFO	Personal Name Series Used For Tracing
UFP	Personal Name AUthor Used For Tracing
UFS	Series Title Used For Tracing
UFT	Topical Subject Used For Tracing
UFU	Uniform Title Heading Used For Tracing

"Refer From" (RF) Tracings:

RFB	Corporate Name Series Refer From Tracing
RFC	Corporate Name Author Refer From Tracing
RFG	Geographic Name Refer From Tracing
RFL	Conference/Meeting Name Series Refer From Tracing
RFM	Conference/Meeting Name Author Refer From Tracing
RFO	Personal Name Series Refer From Tracing
RFP	Personal Name Author Refer From Tracing
RFS	Series Title Refer From Tracing
RFT	Topical Subject Refer From Tracing
RFU	Uniform Title Heading Refer From Tracing

Notes:

NOA	General Reference See Also Note
NOE	General Reference See Note
NON	General Reference See Note
NOR	Catalog Use Note
NOS	Scope Note
NOV	Verification Note

2.1.3.3 Reciprocals

In WLN, cross references and tracings are called "reciprocals" of each other. In Section 2.1.2, former name and later name cross references were also called "reciprocals" of each other. "Reciprocals" means "opposites" in the WLN software. As an authority record is input or updated via Input/Edit, the cross references and tracings are identified and compared against the Authority File. If the cross reference or tracing sortkey matches an authority heading sortkey in the Authority File, the Authority File record is linked to the incoming record, and the text is removed from the incoming authority record's cross reference or tracing field. If the cross reference or tracing sortkey does not match an authority heading sortkey in the Authority File, a new authority record is automatically generated, with the heading text taken from the cross reference or tracing text.

For example, if the following authority record is keyed by an operator:

```
SUT-L   ǂa   ǂJazz music.
SAT-L   ǂa   ǂRagtime music.
```

and "Ragtime music" does not appear in the Authority File, the system automatically creates the following authority record:

```
SUT-L   ǂa   ǂRagtime music.
RFT-L   ǂa   ǂJazz music.
```

"See" cross references and "used for" tracings are reciprocals of each other; "see also" cross references and "refer from" tracings are reciprocals of each other; and "former name" and "later name" cross references are reciprocals of each other. The record that is automatically generated by this reciprocal action is often called a "reciprocal record" or a "generated record".

If a record already exists, the reciprocal heading is merely added to the record. For example, if the following authority record is keyed by an operator:

```
SUT-L   ǂa   ǂAfro-American music.
SAT-L   ǂa   ǂJazz music.
```

the system automatically adds the reciprocal RFT to the authority record:

```
SUT-L   ǂa   ǂJazz music.
SAT-L   ǂa   ǂRagtime music.
RFT-L   ǂa   ǂAfro-American music.
```

f s psychology PF2

COLLECTION ID.ALL BIBLIOGRAPHIC DISPLAY

10. Adult MBD Conference, Psychiatric aspects of minimal brain dysfunction in adults / c1979. xi, 208 p. 79-002585
6. Blumenthal, H. J. Plotinus' psychology; his doctrines of the embodied soul, 1971. 157 p. 75-859494
3. Burton, Asa, Essays on some of the first principles of metaphysicks, ethicks, and theology. 1973. xvi, 414 p. 73-004639
4. Condillac, Étienne Bonnot de, An essay on the origin of human knowledge. 1974. liv, 339 p. 74-147960
8. Husserl, Edmund, The crisis of European sciences and transcendental phenomenology; an introduction to phenomenological philosophy. 1970. xliii, 405 p. 77-082511
9. International Congress on Twin Studies, Twin research : proceedings of the Second International Congress on Twin Studies, August 29-September 1, 1977, . . . 1978. v. 78-008480
13. Main entry Title statement. Edition. Collation. bl 01-000001
1. Moore, Francis Charles Timothy. The psychology of Maine de Biran, 1970. [7], 228 p. 70-468642
5. Pollitt, Jerry Jordan. Art and experience in classical Greece 1972. xiv, 205 p. 74-160094
7. Robinson, T. M. Plato's psychology [1970] ix, 202 p. 76-465044

```
t s psychology

     COLLECTION ID.ALL                                    AUTHORITY DISPLAY
       1. Maine de Biran, Pierre, 1766-1824--Psychology.
       2. Plato--Psychology.
       3. Art--Psychology.
       4. Librarians--Psychology.
 +     5. Psychology.
       6.    --Addresses, essays, lectures.
       7.    --Early works to 1850.
       8. Sex--Psychology.
       9. Twins--Psychology--Congresses.
 M    10. Chronic Disease--psychology--nursing texts
 M    11. Disease--psychology
 M    12. Handicapped--psychology--nursing texts
 M    13. Minimal brain dysfunction--Psychology--Congresses.
 M    14. Psychology--history
```

Figure W-5

```
t s psychology, clinical

    COLLECTION ID.ALL                                AUTHORITY DISPLAY
 *    1. Psychology, Clinical.
M     2. Psychology, clinical
```

Figure W-6

```
b st psychology

    COLLECTION ID.ALL                                    AUTHORITY DISPLAY
+     1. Psychology.
      2.    --Addresses, essays, lectures.
      3.    --Early works to 1850.
+     4. Psychology, Applied.
*     5. Psychology, Clinical.
      6. Psychology, Industrial.
      7. Psychology, Pathological.
      8.    --Addresses, essays, lectures.
      9.    --Congresses.
     10.    --Etiology.
     11. Psychology, Physiological--Collected works.
     12.    --Congresses.
     13. Psychometrics.
     14. Psychotherapy.
     15.    --Congresses.
     16. Psychotherapy patients--Case studies.
     17. Psychotherapy research.
     18. Public health--Brazil.
     19.    --United States.
     20. Public health personnel--United States--Biography.
```

Figure W-7

```
b stm psychology

     COLLECTION ID.ALL                                  AUTHORITY DISPLAY
M      1. Psychology--history
M      2. Psychology, Industrial--essays
M      3. Psychology, clinical
M      4. Psychopathology
M      5.    --Collected works.
M      6. Psychopharmacology
M      7. Psychophysiology--periodicals.
M      8. Psychosomatic Medicine
M      9.    --congresses
M     10. Psychotherapy
M     11.    --congresses
M     12.    --Group
M     13.    --methods
M     14.    --popular works
M     15. Public Health--U. S.--directories
M     16. Radiation Dosage
M     17.    --congresses
M     18. Radiation Protection
M     19. Radioisotopes--diagnostic use
M     20.    --therapeutic use
```

Sample Authority Record in Each of the Display Formats

Complete Display ($,c)

AUTHORITY DISPLAY

```
COLLECTION ID:  1
v1   voc00-12194    db    09/28/81   11/07/83   09/13/84  ----  DLC
 SUT -L  ǂa      ǂAntiques.
 UFT -L  ǂa      ǂAntique collecting.
 UFT -L  ǂax     ǂAntiquesǂCollectors and collecting.
 SAT -L  ǂa      ǂAntiquarian booksellers.
 SAT -L  ǂa      ǂAntique dealers.
 SAT -L  ǂa      ǂArt objects.
 RFT -L  ǂa      ǂArt.
 RFT -L  ǂa      ǂArt objects.
 RFT -L  ǂa      ǂCollectors and collecting.
 NOS     ǂa      ǂHere are entered works on old decorative or
                 utilitarian objects having aesthetic, historic
                 and financial value.  Works on decorative or
                 utilitarian objects are entered under Art objects.
 NOA     ǂa      ǂSee also particular kinds of antique objects,
                 especially the subdivisions Catalogs, Collectors and
                 collecting or Exhibitions when they occur under such
                 objects, e.g. Kitchen utensils; Furniture--
                 Exhibitions.
```

Full Display ($,f)

AUTHORITY DISPLAY

```
     COLLECTION ID. 1
Antiques.
     Here are entered works on old decorative or utilitarian
      objects having aesthetic, historic and financial value. Works
      on decorative or utilitarian objects are entered under Art
      objects.
     See also particular kinds of antique objects, especially the
      subdivisions Catalogs, Collectors and collecting or
      Exhibitions when they occur under such objects, e.g.
      Kitchen utensils; Furniture--Exhibitions.
SA   Antiquarian booksellers.
     Antique dealers.
     Art objects.
```

Headings Display ($,h)

AUTHORITY DISPLAY

```
     COLLECTION ID. 1
+      1.  Antiques.
```

Summary Display ($,s)

SUMMARY DISPLAY

```
     Result:  1 heading.
```

Figure 18

7.3 Expanding on Authority Cross References and Tracings

Authority records can contain cross references which expand or narrow the searcher's original search query. The system allows the searcher to request the "see", "see also", "former name", and "later name" cross references to appear in an authority result set for subsequent searching. The EXPAND command is used to enhance the original search query, without requiring additional keying or Boolean searching by the searcher. For example:

```
T STA antiques $,s
```

results in the display:

```
Result:   6 headings.                          SUMMARY DISPLAY
```

Issuing the SELECT command "S 5 $,f" results in the display:

```
S 5 $,f
                                                AUTHORITY DISPLAY
     COLLECTION ID. 1
Antiques.
     Here are entered works on old decorative or utilitarian
      objects having aesthetic, historic and financial value. Works
      on decorative or utilitarian objects are entered under Art
      objects.
     See also particular kinds of antique objects, especially the
      subdivisions Catalogs, Collectors and collecting and
      or Exhibitions when they occur under such objects, e.g.
      Kitchen utensils; Furniture--Exhibitions.
SA   Antiquarian booksellers.
     Antique dealers.
     Art objects.
```

using the EXPAND command:

```
EXP 5
```

results in the new authority result set:

```
EXP 5

     COLLECTION ID.  1                  AUTHORITY DISPLAY
       1. Antiquarian booksellers.
       2. Antique dealers.
+      3. Antiques.
+      4. Art objects.
```

Occasionally the searcher wants to expand the search to the "used for" and "refer from" tracings, as well as the cross references. The EXPANDALL command is used for this type of expansion.

For example, if a complete display of authority record number 5 from the previous example were requested, all cross references and tracings would appear:

```
S 5 $,c
                                                    AUTHORITY DISPLAY
COLLECTION ID:  1
v1   voc00-12194    db    09/28/81   11/07/83   09/13/84   ----  DLC
 SUT -L  ǂa      ǂAntiques.
 UFT -L  ǂa      ǂAntique collecting.
 UFT -L  ǂax     ǂAntiquesǂCollectors and collecting.
 SAT -L  ǂa      ǂAntiquarian booksellers.
 SAT -L  ǂa      ǂAntique dealers.
 SAT -L  ǂa      ǂArt objects.
 RFT -L  ǂa      ǂArt.
 RFT -L  ǂa      ǂArt objects.
 RFT -L  ǂa      ǂCollectors and collecting.
 NOS     ǂa      ǂHere are entered works on old decorative or
                 utilitarian objects having aesthetic, historic
                 and financial value.  Works on decorative or
                 utilitarian objects are entered under Art objects.
 NOA     ǂa      ǂSee also particular kinds of antique objects,
                 especially the subdivisions Catalogs, Collectors and
                 collecting or Exhibitions when they occur under such
                 objects, e.g. Kitchen utensils; Furniture--
Exhibitions.
```

using the EXPANDALL command:

EXPA 5 or EXL 5

results in the new authority result set:

```
EXPA 5

COLLECTION ID.  1                          AUTHORITY DISPLAY
    3. Antiquarian booksellers.
*   9. Antique collecting.
    4. Antique dealers.
+  11. Antiques.
*  10.    --Collectors and collecting.
+   6. Antiquities.
+   1. Art.
+   7. Art industries and trade.
+   5. Art objects.
+   2. Collectors and collecting.
+   8. Decoration and ornament.
```

3.4 DOBIS as Implemented by the National Library of Canada

3.4.1 Background

DOBIS (Dortmunder Bibliotheks system) is an online library management system originally developed for the University of Dortmund in Germany. In 1977, the National Library of Canada (NLC) and the Canada Institute for Scientific and Technical Information (CISTI) began a project to adapt DOBIS for Canadian standards and requirements and for use in a multi-library network. DOBIS now supports bi-lingual union cataloging for NLC, CISTI, the Library of Parliament, and a number of other Canadian federal libraries. The DOBIS database of over 3.5 million unique bibliographic records is available to its user libraries and to libraries subscribing to its Search Service. While DOBIS searching is described in NLC brochures as "user-friendly," the system is used largely by trained librarians rather than end-users. The DOBIS Search Service provides two days of intensive training for designated coordinators, and only in a few specialized libraries is the system used as an online catalog.

3.4.2 Description of the Subject Authority File

The NLC/DOBIS system supports the creation of both bibliographic and authority records in a MARC-compatible format. Authority records can be created online for names, titles, and subjects. Subject authority records are coded to indicate the source list of the heading. A sample subject authority record is shown in Figure D-1.

All bibliographic and authority records are indexed together by the access point files (APF). There are separate APF's for Names, Titles, Subjects, Classification, and a variety of control numbers. The APF records for Names, Titles, and Subjects themselves contain sufficient linking, cross reference, and source information to serve as authority control records. In fact, NLC/DOBIS users do not create and maintain full subject authority records, but instead rely on the Subjects Access Point File. A complete list of the fields contained in a Subject APF record is shown in Figure D-2. These fields include: subject type (topical, geographical, etc.), subject source (e.g., LCSH), language (French or English), links to authority and bibliographic records (i.e., record i.d. numbers and counts of the number of linked records), and links to other APF headings. APF headings may be keyed-in online (as would be the case with cross-references) and are generated automatically when bibliographic records are keyed or loaded.

There is no limit to the number of subject list sources from which subject authority headings or subject access points may be drawn. At present, the following sources are coded separately in NLC/DOBIS: LCSH, Annotated Card Program, MeSH, NAL, Canadian Subject Headings (CSH), Repertoire de vedettes-matiere (RVM, a translation of LCSH supplied by Universite

Laval), NASA, Canadian Urban Thesaurus, and PRECIS. Of these, LCSH, CSH, and RVM are the lists most commonly used by NLC/DOBIS libraries. Codes for subject source list in the Subject APF are derived from the appropriate indicator and subfield values in the 6xx fields of bibliographic records. Different subject headings on a particular bibliographic record need not all be derived from the same source list.

Subject source and language are included in the sort form of a subject headings. Thus the same text string coded differently for source and/or language will generate a separate subject file record. This permits, for example, a different cross reference structure to be created for a term as it is used in different source lists. Because there is considerable overlap between LCSH and CSH, a search, for example, on the subject "Canada-Foreign relations-United States" might produce the following result:

Canada-Foreign Relations-United States	CSH 100
Canada-Foreign Relations-United States	LCSH 100,

requiring the searcher to request two separate files of retrieved documents. Catalogers are trained not to create duplicate subject file records online, but to link headings to an existing APF record to reduce such duplicative results.

3.4.3 Links and references

Since the access point files serve as indexes to DOBIS bibliographic and authority files, Subject File records include the record numbers of the bibliographic and authority records to which they point. The link is reciprocal--APF record numbers are likewise contained in bibliographic records. Only complete headings are linked; there is no provision for linking only part of a heading from a bibliographic record (e.g., a main heading alone or a geographic subdivision) to an access point record. This link permits headings maintenance through the APF, since a change in to a Subjects File record automatically changes the heading in all linked records. Since separate APF records are maintained for subjects from different sources, headings from a variety of thesauri can be maintained on the system.

Within the Subjects Access Point File a variety of links and references can be specified, including hierarchical relationships. Figure D-3 shows the types of linkages available. Links and cross references to a heading can only be added after the heading has been established in the APF. Subject see references (i.e., non-postable terms) must also be coded for source and language as are main terms. Pointers between Subjects File records are reciprocal; the user need only create the link in one direction and the system will automatically supply the appropriate reciprocal link. Links can be made between headings regardless of subject source, e.g., a CSH term may be related to an LCSH term. A common cross-thesaurus link is the French/English equivalence relationship. Links provide assistance in searching as described in the next section.

3.4.4 Retrieval and display

In DOBIS, subject searching is defined as a search of the Subject APF. Each search term is searched as a left match with implicit right trunication. All searches lead to an alphabetical display of the Subject File, even when there is one term that precisely matches the search term entered. Figure D-4 shows a sample Subject File display.

It is not possible to limit searches by subject source even though this information is carried in the index. Subjects from all sources are displayed in an integrated alphabetical list, with the subject source clearly indicated. The number of bibliographic records linked to each heading is also shown, including headings representing non-postable terms. Searchers may select a heading and move directly into a summary display of bibliographic records linked to that heading. Thus a term which is postable in one thesaurus and non-postable in another would appear twice, as would identical terms coded for different source lists. For example:

x	Canada-History-War of 1812	LCSH	10
	Canada-History-War of 1812	CSH	50
	Canada-History-War of 1812	CUT	2

Terms that are cross-references or that are linked to other Subject File headings are identified by an "x." Searchers can view related terms (i.e., those linked to the search term) by requesting the appropriate display. Figures D-5 - D-7 show a detail display of the APF record for the term "Canada-Foreign relations-United States" and a summary and detail display of another APF term linked to it--in this case its French equivalent from Repertoire de vedettes-matiere.

Multiple thesauri have not posed a problem in DOBIS retrieval, largely because: 1) most DOBIS searchers are trained specialists, and 2) the most commonly used thesauri are LCSH, CSH, and RVM, which are highly compatible. (CSH is essentially a list of exceptions and additions to LCSH, while RVM is a translation.) Both the alphabetical Subjects File display and the displays of linked headings allow knowledgeable searchers to browse, refine, and select subject terms from a variety of subject sources.

Figure D-1

Subject Authority Record

Full Encoding Level

```
Cataloguing authorities search
Authority number
Full information    Authority      100        LC          verified
                                              accepted - subject  LCSH

SUBJECT:
  Topical LCSH  e vn: $aColonization
NOTES:
  cat src   n.a. : $aDLC$beng$cCaOONL
  scope/s   n.a. : $aHere are entered works on the policy of settling inhabitants
                   or nationals in colonial areas.  Works on migration from one
                   country to another are entered under Emigration and immigrat
                   ion.  Works on migration within a country are entered under M
                   igration, Internal.  Works on general colonial policy, includ
                   ing land settlement, are entered under Colonies.
  gen s.a   n.a. : $isubdivision$aColonization$iunder names of countries, region
                   s, etc., e.g.$aAfrica--Colonization$iand under names of class
                   es of persons, e.g.$aAfro-Americans--Colonization; Epithetic
```

```
Cataloguing authorities search
Authority number
Full information    Authority      100        LC          verified
                                              accepted - subject  LCSH

NOTES (cont.):
                   ses of persons, e.g.$aAfro-Americans--Colonization; Epithetic
                   s--colonization; Mennonites--Colonization
  se/note   note : $aAgricultural colonies
  fnd/s     n.a. : $aLCSH7
```

```
Cataloguing authorities search
Authority number
Full information -  authority codes

Heading type                subject
Encoding level              full
Record status               new
Record type                 accepted
Status of heading           usable
Status of references        n.a.
Mod. record ind.            mod
Cataloguing source          LC
Lang. of catalog.           English
Catalog. date               830614
Date last trans.            830614

        Input source        On-line
```

```
Cataloguing authorities search
Authority number
Full information -  authority codes

Subject control information
  Direct/indirect descriptor    1 subdiv
  Romanization scheme           not rom.
  Language romanized            nnn
```

Subject Authority Record

Incomplete Encoding Level

```
Cataloguing authorities search
Authority number
Full information    Authority      99         LC          verified
                                              accepted - subject  LCSH

SUBJECT:
  Topical LCSH  e vn: $aColonization
NOTES:
  cat src   n.a. : $aDLC$beng$cCaOONL
  fnd/s     n.a. : $aLCSH7
```

```
Cataloguing authorities search
Authority number
Full information -  authority codes

Heading type                subject
Encoding level              incoml
Record status               new
Record type                 accepted
Status of heading           usable
Status of references        n.a.
Mod. record ind.            mod
Cataloguing source          LC
Lang. of catalog.           English
Catalog. date               830614
Date last trans.            830614

        Input source        On-line
```

```
Cataloguing authorities search
Authority number
Full information -  authority codes

Subject control information
  Direct/indirect descriptor    1 subdiv
  Romanization scheme           not rom.
  Language romanized            nnn
```

SUBJECT ACCESS POINT FILE (AFSU,3)

```
/*****************************************************************/C01
/*  CHANGE DESIGNATOR:  C01                                      */C01
/*                                                               */C01
/*  REASON:  AUTHORITY ISOMORPHISM                               */C01
/*                                                               */C01
/*  CHANGES:  1. APF MAP BIT 5 IS NOW UNUSED.  IT PREVIOUSLY WAS */C01
/*               USED TO INDICATE PRESENCE OF AUTHORITY NOTES WHICH */C01
/*               WILL BE STORED IN THE AUTHORITY RECORD.         */C01
/*            2. APF MAP BIT 9 HAS BEEN REDEFINED. AN AUTHORITY LIST */C01
/*               INSTEAD OF AUTHORITY DATA WILL BE IN APF RECORDS. */C01
/*            3. THE APF LINKAGE HAS BEEN EXPANDED TO CONTAIN    */C01
/*               TRACING INFORMATION AND REFERENCE GENERATION CODE. */C01
/*            4. THE APF LINKAGE CHANGE-IN-NAME INDICATOR HAS BEEN */C01
/*               RENAMED 'SEE ALSO RELATOR'.                     */C01
/*                                                               */C01
/*  DONE BY:  E. PEDERSON                DATE:  NOVEMBER 2, 1982 */C01
/*****************************************************************/C01

/*****************************************************************/C02
/*  CHANGE DESIGNATOR:  C02              TASK NO.:   W1633        */C02
/*                                                               */C02
/*  REASON:       IMPLEMENTATION OF 2ND EDITION OF CAN/MARC      */C02
/*                AUTHORITY FORMAT                               */C02
/*                                                               */C02
/*  CHANGES:  IN APF LINKAGES:                                   */C02
/*               1. RENAME 'TRACING INFORMATION' TO 'PRINT CONSTANT */C02
/*                  CODE'.                                       */C02
/*               2. DELETE 'SEE ALSO RELATOR'.                   */C02
/*               3. ADD 'EARLIER CATALOGUING RULES' AND 'FORMERLY */C02
/*                  ESTABLISHED HEADING' CODES.                  */C02
/*                                                               */C02
/*  DONE BY:  E. PEDERSON                DATE:  JUNE 11, 1985    */C02
/*                                                               */C02
/*****************************************************************/C02
```

PRIMARY ENTRY

OFFSET	LENGTH	TYPE	DESCRIPTION	
0	2	BIT	STANDARD APF MAP	
		BIT 1	UNUSED	
		BIT 2	1 IF DISPLAY FORM PRESENT	
		BIT 3	UNUSED	
		BIT 4	1 IF LINKAGES PRESENT	
		BIT 5	UNUSED	C01
		BIT 6	UNUSED	
		BIT 7	1 IF DOCUMENT LIST PRESENT	
		BIT 8	UNUSED	
		BIT 9	1 IF AUTHORITY LIST PRESENT	C01
		BIT 10	1 IF LANGUAGE EQUIVALENCES PRESENT	
		BIT 11	1 IF 'SEE' REFERENCE PRESENT	
		BIT 12	UNUSED	
		BIT 13	UNUSED	

			BIT 14 UNUSED	
			BIT 15 UNUSED	
			BIT 16 UNUSED	
2	1	CODE	SUBJECT TYPE	
3	1	CODE	SUBJECT SOURCE	
4	1	CODE	LANGUAGE OF APF	X
5	1	CODE	VERIFICATION LEVEL	
6	5		FILLER (ALWAYS HEX ZERO)	
11	2	BINARY	LENGTH OF SORT FORM OF SUBJECT	
13	VAR	CHAR	SORT FORM OF SUBJECT	
- - - - - - - - - - - - - - - -			ALWAYS PRESENT UP TO HERE - - - - - - - - - -	

2	BINARY	LENGTH OF DISPLAY FORM OF SUBJECT	
VAR	CHAR	DISPLAY FORM OF SUBJECT INCLUDING SUBFIELDS	
		AUTHORITY NUMBER LIST	C01 C01
4	BINARY	AUTHORITY DOBIS NUMBER	C01
2	BINARY	COUNT OF AUTHORITY DOBIS NUMBERS	C01
		APF LINKAGES	
15		LINKAGE	
1	CODE	APF LINKAGE TYPE	
5	POINTER	APF LINKAGE POINTER	
4	BINARY	DATE OF LAST TRANSACTION	
1	CODE	STATUS	
1	CODE	PRINT CONSTANT	C02
1	CODE	REFERENCE GENERATION CODE	C01
1	CODE	FORMERLY ESTABLISHED HEADING	C02
1	CODE	EARLIER CATALOGUING RULES	C02
2	BINARY	NUMBER OF LINKAGES	
		DOCUMENT NUMBER LIST	
4	BINARY	DOCUMENT NUMBER	
2	BINARY	COUNT OF DOCUMENT NUMBERS	

Figure D-3

Cataloguing Cross-References Screenflow

```
*** SERIES/1 '  ITS   ' - DISPLAY PRINT FOR TERMINAL  6   07:18:03     03    ***


 Catalogue search
 Subjects
 Linkage type

   1 not specified                    7 see also:
   2 linkage not used                 8 see also from:
                                      9 dual s.a/s.a from:
   3 hierarchy to:
   4 hierarchy from:                 10 English equivalence:
                                     11 French equivalence:
   5 see:
   6 seen from:                      Valid/Invalid Numbers
                                     12 corresponds to valid/correct no.
                                     13 corresponds to invalid/incorrect no.



  Enter number
```

Figure D-4

```
*** SERIES/1 '  ITS    ' - DISPLAY PRINT FOR TERMINAL  6    06:48:42      03     ***

Searching
Subjects

                                                                                  LCSH
 1      CANADA FOREIGN RELATIONS TREATIES UNITED STATES                          LCSH      1
 2*X    Canada Foreign relations United States                                   LCSH    258
 3      Canada Foreign relations United States Addresses, essays, lec            LCSH   / 11
 4      Canada Foreign relations United States Bibliography                      LCSH      2
 5      Canada Foreign relations United States Congresses                        LCSH     10
 6      Canada Foreign relations United States Periodicals                       LCSH      2
 7      Canada Foreign relations United States Sources                           LCSH      3
 8      CANADA FOREIGN RELATIONS VIETNAM                                         LCSH      1
 9      CANADA FOREIGN RELATIONS WEST INDIES, BRITISH                            LCSH      1
10      Canada Foreign relations Yearbooks                                       LCSH      1
11      Canada Foreign relations 1867-  Addresses, essays, lectures             CSH       1
12      Canada Foreign relations 1867-1918                                       LCSH      1
13      CANADA FOREIGN RELATIONS 1918-1930                                       LCSH      1
14      Canada Foreign relations 1918-1945                                       LCSH      4

Enter number or code
/d/2
t new term    f forward
i new file    b backwrd
```

indicator that this heading is used in a cross-reference linkage

Subject Source Indicator

Count of Bibliographic Records with in which Subject Heading is used.

d detail a add

↑ command to display screen on which cross-references can be viewed.

Figure D-5

```
*** SERIES/1 '  ITS   ' - DISPLAY PRINT FOR TERMINAL  6   06:50:09     03     ***

Searching
Subjects
                        Detail information

 Canada Foreign relations United States


 1 Documents                  258        Function / source      LCSH
 2 Subfield codes                        Type                   geograp
 3 Cross references             1
 4 Authorities                  0
                                         Lang. of access point  English
Language equivalences present            Verification level     authenticat

Enter number or code
3

w show file
```

Figure D-6

```
*** SERIES/1 '  ITS   ' - DISPLAY PRINT FOR TERMINAL  6   06:53:03     03    ***


Searching
Subjects
                        Summary

 Canada Foreign relations United States

                                                                    Documents
  1 fre equiv          ¬ former      ¬e.rules      g fr lin      n.a.        112
    Canada Relations ext'erieures 'Etats-Unis
```

↑ related heading (in this case, the equivalent French form of the English heading)

↑ Count of bibliographic records in which the related heading is used.

```
Enter number or code
1

                                                          e end
```

Figure D-7

```
*** SERIES/1 '  ITS   ' - DISPLAY PRINT FOR TERMINAL  6   06:53:21     03    ***


 Searching
 Subjects
                        Detail information

  Canada Relations ext'erieures 'Etats-Unis




   1 Documents                  112          Function / source       RVM
   2 Subfield codes                          Type                    geograp
   3 Cross references             1
   4 Authorities                  0
                                             Lang. of access point   French
 Language equivalences present               Verification level      authenticat

 Enter number or code


                                                          e end
```

3.5 UTLAS

3.5.1 Background

UTLAS International, U.S. is a large computer service organization that provides, along with other services, online support for cataloging, bibliographic file maintenance, and authority control for customer libraries in North America and Japan. Originally designed as a cataloging system for the University of Toronto Libraries, UTLAS began providing service for other Canadian libraries in 1973. Incorporated in 1983, UTLAS was sold by the University in 1985; it is now a wholly-owned subsidiary of International Thomson.

The UTLAS Catalogue Support Systems (CATSS) allows customers to create and maintain their own separate bibliographic databases on the UTLAS computer, deriving records from copy present in the database or entering original cataloging. UTLAS also loads the MARC bibliographic files of the National Library of Canada, the Library of Congress, and the British Library, along with other bibliographic and authority files, so that its combined databases contain over 24 million records. The CATSS system is designed to support technical processing, reference and verification searching, and ILL; at present, CATSS does not provide retrieval results designed for library patron searching.

3.5.2 Description of the Subject Authority Files

The CATSS system supports a large number of logically separate authority files that can be linked to headings in customers' bibliographic databases. The company has maintained a source authority file of Library of Congress Subject Headings, created from the 8th edition and updated online from printed additions and changes. It plans to load the latest cumulation and process weekly updates. Additional LCSH headings--e.g., changed pattern headings or strings that combine headings and subdivisions authorized by example but not enumerated in LCSH--are sometimes added to the LCSH authority file based on notices in the LC Cataloging Service Bulletin. In addition, customers maintain through online input a number of private subject authority files, including the 70,000-record Repertoire de vedettes-matiere (RVM), a French translation of LCSH created and maintained by Universite Laval. Customers' private subject files range from local variations on LCSH to extensive files of PRECIS strings. The files are distinguished by ranges of sequence numbers, with associated codes for subject source agency. A customer would need to define two private files in order to maintain separate authority files for two different thesauri. An identical term appropriate to both files would be repeated in each, with separate reference structures.

UTLAS supports most fields in the MARC Authorities Format, adding a few fields for local control and for French/English language verification (allowing a customer to create a private

file that contains both languages, yet verify appropriate headings against one language or the other). A sample UTLAS authority record is shown in Figure U-1. Authority records may be loaded from MARC format source files (e.g., the LC Name Authority File), derived by editing existing records online, or keyed in online; authority records are not updated through loads of bibliographic records.

Headings on bibliographic records are verified against the authority files (either during online creation or loading) according to profiles developed for each customer. Profiling for subject validation takes into account the MARC tag of the heading field, the second indicator value (currently only used to indicate whether or not the heading should match to the LCSH file), and the customer's selection of subject authority files in determining matches. Coding in the $w subfield can also be regarded to define links. A customer selects valid subject authority files in priority order, e.g., LCSH may be a customer's first choice, with "no matches" to the LC file verified against a second choice local file. The process of validation against a series of authority files is known as "cascading," and is particularly useful when one authority file (e.g., a customer's local file) is essentially a list of variations, exceptions, or additions to a larger file. A customer may develop different profiles for subfiles within its database (e.g., branch libraries). The verification comparison is used to identify headings which are new to the authority file (or possibly erroneous) and to link matched headings for future maintenance. UTLAS can also provide a "second pass" on unlinked subject headings, matching only the "a" subfield for validation.

3.5.3 Links and references

In the validation process, subject authority records are linked to headings in bibliographic records and to other authority records to aid in file maintenance. Each authority record is assigned a unique number, or ASN. As bibliographic record headings are validated against an authority file, if an exact match to a 1xx or a 4xx of an authority record is found, the bibliographic heading is replaced with the ASN of the authority record. At the same time, the bibliographic record number (RSN) is added to an index of matching bibliographic and authority records. Any subsequent changes to the 1xx field of the authority record will result in an automatic correction to the linked bibliographic records. Bibliographic headings are linked to selected authority files based on tags, indicators, and a library's profile. Each specific link is made only once, to one authority record.

It is also possible to substitute an ASN for only the matching "a" subfield of a topical or geographic heading; this link can be done automatically on a "second pass" verification of loaded bibliographic records, or can be requested online when a bibliographic record is created. Another online command, "ALINK" will allow an online user to create a link from a subfield

in a bibliographic heading (e.g., a geographic subfield) to an authority record, thus permitting maintenance directly on subdivisions.

In much the same way that bibliographic and authority records link to each other, it is possible to link UTLAS authority records to one another, a procedure known as "nesting." The most obvious use of this link is for subject headings with subdivisions. For example, an authority 150 might consist of an ASN for a main term followed by the text of a subdivision. Alternatively, if the subdivision is commonly used, it too might have a separate authority record and ASN; then the 150 for the entire string would consist of two ASN's--one for the main term and one for the subdivision. An example showing the creation of a nested authority record for the term "Phonetics-Congresses" is shown in Figure U-2. This authority-to-authority link can be built across authority files, e.g., to create and maintain a local heading that consists of an LC main term and a local subdivision.

While UTLAS customers use this technique in private files, UTLAS has not created authority records for free-floating subdivisions in its LCSH source file. However, UTLAS plans to extend its "NEST" feature to ease subdivision verification and maintenance. The NEST process will be introduced to bring under authority control free-floating, pattern, chronological, and geographical subdivisions for tags 650 and 651. Authority subdivision records will be created by UTLAS staff. These records will be clearly identified as NEST records and will indicate subdivision type. Enough information will be present for the system to determine which subfield ($x, $y, $z) to match. For subfield $z, NEST will first attempt to link all $z's to a single nest record. Failing that, any remaining unlinked $z subfields will then be validated singly. $x and $y subfields will always be validated singly. Examples are shown in figure U-4.

In conformance to LC practice, UTLAS subject authority records can include 4xx (see from) and 5xx (see also from) fields. An online command, "VALI", permits users to validate 5xx fields in an authority record, comparing each to the 1xx in other authority records. The 5xx fields are then linked to the appropriate ASN, so that a change to the linked main term will correct the 5xx reference. Choice of authority files for such validation and linking is based on the customer's profile; it is not necessarily restricted to within a single authority file.

A special command has been developed within UTLAS to provide for validation and translation of bilingual or French users' LCSH terms into their French equivalent in RVM. Records in the RVM authority file contain two 150 fields; the first is the authorized RVM heading and the second is the LCSH term from which

it was translated. This device enables the automatic translation and linking of French and English headings without creating a 4xx from the English version.

3.5.4 Retrieval and display

As noted earlier, UTLAS searching and displays are designed to support technical processing. The results supply considerable useful information to catalogers, but are not end-user friendly. UTLAS is planning the development of an online catalog system that will combine many of the powerful features of UTLAS authority control with improved searching.

Subject searches on UTLAS are exact string matches with right truncation possible after seven characters. The indexes to the UTLAS authority and bibliographic files may be searched separately or together. When a search produces a hit in both files, the authority record is listed first, as shown below in a sample subject search on "Dogs":

	File	Authority	ASN/RSN	Brief display	Tag of heading in record
1	USO	A	0066745016	Dogs	150
2	USO		0066743251	Dog stories/ Jones	650

Searches of the subject authority file will retrieve hits from all authority files included in the user's profile; the retrieval from each authority file will be sorted separately, in order according to the users' profiled validation heirarchy. A sample search on the term "Philology" is shown is Figure U-3. The figure requires some explication. The results display the 1xx field of each authority record in which the search term appears. To the right of the result is an indication of the field tag assigned to the search term, in this case "Philology," in the retrieved authority record. Thus, "Philology" is included as a 150 in the RVM authority record for "Philologie" (an example of the double 150 used for RVM headings and their LCSH originals) and is a 550 reference on the LCSH record (the SAF file) for "Archaeology." Unfortunately, within each authority file sequence the records are sorted by control number, leaving the appearance of "Philology" as a 150 mixed in with the references. Matches to 450 fields in authority records are displayed in the same manner as the 550 references. Once understood, the display enables the cataloger to view related terms from a number of subject lists.

AUTHORITIES TRAINING GUIDE

1. INTRODUCTION

The UTLAS database is divided into two basic sections: bibliographic records, which contain full cataloguing information for both print and non-print items; and authority records, which contain the preferred form, see and see also references for personal, corporate and conference entries, uniform titles, series and subject headings.

Sample Authority Record:

```
     ASN     PTC  OPN    DFC      DCR      DCH   TCH   LAN
   61478052 updt EPLA 79Jun07 73Jul06 820ct20 19:29 eng
STATUS           fin
                 01:730706 02:          i 12:      a 13:     013 19:     d
                 30:     y
010  .......    0001 .750101$ash 78106682
040  .......    0001 .750101$aDLC$dae
U040 .......    0001 .750100$aae$beng
053  .......    0002 .750101$aG104-8
150  .0.....    0001 .750101$w1a3a$aNames, Geographical
450  .0.....    0002 .750101$w1a3a$aGeographical names
450  .0.....    0003 .750101$w1a3a$aPlace-names
550  .0.....    0011 .750101$w1a3a$aGeography$xTerminology
550  .0.....    0012 .750101$w1a3a$aNames
550  .0.....    0013 .750101$w1a3a$aToponymy
550  .0.....    0014 .821020$w1a3a$aOnomastics
A61-478-052 Cancelled 28-Jun-85 Logon 5509,DEMA:DEM
```

```
Authority record for subdivision:

    ASN    PTC  OPN    DFC     DCR     DCH    TCH  LNG
  61586020 orig ALBA 75JanOO 81Nov24 81Nov24 12:25 eng
STATUS           fin
                    01:811124 02:      n 12:     a 19:    e 30:      y
U040 ........  0001 .750100$aaeu$beng
150  .0......  0001 .811124$w$aCongresses
Anymore?


Authority record for main heading:

    ASN    PTC  OPN    DFC     DCR     DCH    TCH  LNG
  61486347 orig LC   79Jun07 73Jul06 80Feb14 00:00
                    01:730706 12:    a 13:   024 19:     d 30:     y
010  ........  0001 .750101$ash 78119812
040  ........  0001 .750101$aDLC
053  ........  0002 .750101$aP221-7
150  .0......  0018 .750101$w1a3a$aPhonetics
360  ........  0018 .750101$w1a3a$isubdivisions$aPhonetics, Phonology,$iand
                    $aPronunciation$iunder names of languages, e.g.$aFrench
                    language--Phonetics; English language--Phonology; German
                    language--Pronunciation
450  .0......  0019 .750101$w1a3a$aOrthoepy
450  .0......  0020 .750101$w1a3a$aPhonology
550  .0......  0021 .750101$w1a3a$aLinguistics
500  .0......  0022 .750101$w1a3a$aSound
550  .0......  0023 .750101$w1a3a$aSpeech
550  .0......  0024 .750101$w1a3a$aVoice
Anymore?
.inse 150,(...)/($xCongresses)
150  .0.....:  2001 .750101$w1a3a$aPhonetics$xCongresses
Anymore?
.ALINK 150 $a,61486347
150  .0......  2001 .750101$w1a3a
                    $aPhonetics((ASN=61486347))
                    $xCongresses
Anymore?
.alink 150 $x,61586020
150  .0......  2001 .750101$w1a3a
                    $aPhonetics((ASN=61486347))$xCongresses((ASN=61586020))
Anymore?
dele 360,450 a,550 a
Done
Done
Done
Done
Done
Done
Done
Anymore?
```

```
What the new heading looks like:

    ASN     PTC  OPN    DFC      DCR      DCH     TCH      SNR    LNG
  99001532 derv DEMA 79Jun07 73Jul06 85Jul09 10:36 61486347 eng
STATUS           fin
                    01:730706 12:    a 13:   024 19:    d 30:    y
040  ........  0001 .850709$aDLC
U040 ........  2001 .850709$aDEM$beng
053  ........  0002 .850709$aP221-7
150  .0......  2001 .850709$w1a3a
                    $aPhonetics((ASN=61486347))$xCongresses((ASN=61586020))
Anymore?



What the validated field looks like:

650  .0......  2001 $aPhonetics$xCongresses.
Anymore?
veri 650
650  .0....0.  2001 $aPhonetics$xCongresses((ASN=99001532,61486347,61586020))
                    VALIDATED : ((99001532)) 150
Anymore?
```

Figure U-3

```
Access key: 6/philoloov:a
Searching
  14 records retrieved
Enter hitlist commands: +14
    1 RVM A 0008018798 Philoloaie                                    150
    2 SHN6A 0061417605 Archaeoloov                                   550
    3 SHN6A 0061453981 Grammar. Comparative and general              550
    4 SHN6A 0061466956 Language and languages                        550
    5 SHN6A 0061469175 Literature                                    550
    6 SHN6A 0061469200 Literature. Comparative                       550
    7 SHN6A 0061486240 Philoloov                                     150
    8 SAF A 0020002174 Archaeoloov                                   550
    9 SAF A 0020033578 Literature                                    550
   10 SAF A 0020033604 Literature. Comparative                       550
   11 SAF A 0020070191 Grammar. Comparative and general             550
   12 SAF A 0020080606                                              550
   13 SAF A 0020088249 Language and languages                       550
   14 SAF A 0020101894 Philoloov                                    150
```

Example 1: $x, $y

Nest authority record for pattern and chronological subdivisions are created and identified as authority subdivision records (type x or y).

```
ASN X
1xx$aHomes and haunts
4xx$aHomes and hovels

RSN
600$aShakespeare, William$xHomes and hovels
veri
no match
verilead
600$aShakespeare, William(ASN)$xHomes and hovels
nest
600$aShakespeare, William(ASN)$xHomes and haunts (ASNX)
```

Example 2: $z : Geographical Subdivisions - Authority Records with Two Parts

```
ASN Z
1xx$aFrance$zParis

RSN
650$aCuisine$zFrance$zParis

veri
no match
verilead
650$aCuisine(ASN)$zFrance$zParis
nest
650$aCuisine(ASN)$zFrance$zParis(ASNZ)
```

Example 3: $z : Geographical Subdivisions - Separate Auth Records

```
ASN Z1                                          ASN Z2
1xx$aGermany (West)                             1xx$aTubingen

RSN
650$aCuisine$zGermany (West)$zTubingen
verilead
650$aCuisine(ASN)$zGermany (West)$zTubingen
nest
650$aCuisine(ASN)$zGermany (West)(ASN Z1)
                 $zTubingen (ASN Z2)
```

:ds

3.6 Geac Bibliographic Processing System

3.6.1 Background

The Geac Bibliographic Processing System (GBPS) is an online catalog management system developed and marketed by Geac Computers International and operated on Geac's Concept 9000 multiprocessor. Cataloging, catalog maintenance, and online public access functions are included. The system is designed to support sophisticated management of bibliographic and authority records, and allows users to define over 40 kinds of linking relationships between records (e.g., preceding/succeeding title links, headings links to authority files, etc.). The system allows users considerable flexibility; many aspects of file definition, linking, verification, and display are table driven, providing a wide variety of options for establishing bibliographic and authority files.

Geac installed a prototype version of GBPS at the Biblioteque Nationale in June 1985; University of Waterloo began testing the field version in September of that year. The first production release of GBPS was installed in April 1986 at New York University and the Smithsonian Institution. Since customers are still in the process of experimenting with the system and defining files and profiles, no site is yet operating a catalog using multiple subject authority files.

3.6.2 Description of the Subject Authority File

The U.S. MARC Format for Authorities is supported, and customers may define other authority record formats as well. Authority records may be loaded from tape, received via an online interface (the GBPS implementation of the LSP protocol is currently in field test), or keyed in online. Optionally, authority records can be created from bibliographic headings as bibliographic records are loaded. In this situation, a customer may elect to have provisional authority records created which will not reside in the core authority file until validated.

A site may create as many authority files as it needs (the upper limit is ca. 8,000), including multiple subject authority files. The definition of authority files can be tailored to each site, since customers are provided with a complex set of options from which to choose. Each customer defines a set of record types that will be used for its collections. These record types may simply be standard MARC formats (e.g., books, scores, films) or they can be more specific (e.g., medical books, exhibition catalogs, ethnographic recordings). For each record type, up to ten different authority files can be specified for each MARC tag; the files are defined by tag and an indicator value. For example, a site could define "books" as a record type. Topical subject headings (i.e., 650 fields) assigned to "books" could be linked to any one of 10 different subject authority files, depending on indicator value. Or, the site could be more specific by defining record types for "general books" and "art

books," with up to ten different subject authority files possible for governing the 650 fields in each type.

Record type also serves to partition the indexes, so that materials of the same type can be searched separately. For purposes of subject authority control, the simplest way to define files would be to use one thesaurus for each record type within a database, permitting separate searching of each thesaurus. If multiple thesauri are used for each record type (which is possible since up to 10 may be defined for each tag), the headings from the different thesauri used will be retrieved together. The headings from the different thesauri could be identified by source--a site could define this display as a trailing print constant. For example, records containing both LCSH and ACP (juvenile) subjects could be provided with authority control for each set of subjects. The two sets of subjects would be retrieved together in the same search, with the source of heading displayed. If segregated subject searching were desired, this could be achieved by creating two bibliographic records (one easily copied from the other) identified as different record types or as belonging to different collections. The record types could then be specified in the search argument, allowing the user to search one subject authority file or all subjects. For example, a subject search could specify record types controlled only by LCSH, could specify those controlled by ACP, or could lack specification and retrieve LCSH, ACP, and subject headings not linked to an authority file.

In addition to record type, index partioning may be based on other qualifiers such as language, date, and location. If the use of a particular thesaurus were confined to sets of records that could be defined by qualifier (e.g., only RVM headings used on French language records, or MeSH only assigned to materials located in the Medical Library), subject searches restricted by qualifier would also be searching only the relevant vocabulary.

3.6.3 Links and references

Headings in bibliographic records are linked to headings in authority records via a common index that carries the linked record numbers; this index supports headings maintenance. As described in the preceding section, links are specified by tag and indicator, so that, for example, geographic subject headings could be linked to a different authority file than topical headings. Links are normally made for an entire heading. However, within any particular file (i.e., a record type within a collection) it would be possible to decide, for example, to define the links for tag 650, second indicator 0, such that all $a subfields (main topic) were linked to one authority file and all $z subfields (geographic subdivisions) were linked to another. The designer of GBPS notes that such separate linking of subdivisions compromises authority control over the whole string. For example, "Potato chips--Personal narratives" would not be identified as a new heading when it entered the system.

An alternative way to gain control over subdivisions would be to create an authority file for subdivisions and link authority records to it.

Within an authority file, records can be linked through related term references. Broader, narrower, and related term relationships can be specified. The system will accept a "blind" related term reference, but will not display that reference in an online catalog search. New terms or related term references that conflict with a "see from" reference (a 450 entry) are reported as errors. Machine validation and linking only occur within a particular authority file and cannot be specified across files. However, through manual input, an editor could specify related term relationships that had been defined across files (e.g., a defined 550 reference in an LCSH record that displayed as "Search also under Medical heading..."). However, such a "see also from" reference could not exist simultaneously with a "see from" reference (450 field) from the same term.

3.6.4 Retrieval and display

Subject headings for authority and bibliographic records are contained within the same index; the authority file can also be searched separately if desired. The subject indexes include a subject string index (searchable by exact or truncated match) and a subject keyword index. Alphabetical displays of both indexes are available; these include cross-references ("rejected terms") and numbers of postings for each heading. In the cataloging module, related terms (550 entries) are included in the index display; in the public online catalog search these may be requested by a follow-on command. Related terms are displayed in alphabetical order by term, although a sort grouped by RT, BT, and NT could be developed if records were adequately coded.

Whether or not terms from different thesauri are retrieved in a particular subject search is dependent on the profiles established at each site. Displays could be developed that distinguished headings and references retrieved by source of list. At present, the only GBPS user engaged in planning for multiple subject authorities is the Smithsonian Institution Library. In its initial implementation, the Smithsonian plans not to mix thesauri within a single collection; within each logical catalog only one controlled subject vocabulary will be used. Users who wish to search more than one catalog will do so sequentially.

3.7 NOTIS

3.7.1 Background

Northwestern Online Total Integrated System (NOTIS) is an integrated, comprehensive library system with operational functional components for database management, cataloging, acquisitions, serials control, public access catalog, authority record management, and circulation. NOTIS is a software package developed and used by Northwestern University Library since 1970. The foundation of the NOTIS system is a file of bibliographic records linked to records in various secondary files and accessed through automatically generated and full-heading index files. The system runs on IBM or IBM-compatible equipment.

The online catalog module of NOTIS is called LUIS (Library User Information System) and was first implemented at Northwestern in January 1980. It is a command driven system that provides for user prompting and help screens. Browsing of index entries is supported. Automatic right-hand truncation is provided.

3.7.2 Description of the subject authority file

NOTIS supports the MARC Format for Authorities. The system provides for the addition and modification of authority records using keyboard, tape, or an online interface to OCLC. Libraries in a NOTIS installation have options for developing the relationship between their bibliographic records and authority files. Institutions sharing one installation can share one authority file while maintaining separate bibliographic records. Alternatively, libraries sharing a system can create separate authority files. It is also possible for a single institution to create authority files for different subject heading lists.

The system allows for the automatic generation of authorized headings files, global changes, lists of headings new to the database, and lists of headings deleted from the database. One of the system's newest capabilities is the automatic generation of cross-references from the authority records for the online catalog indexes. Libraries also have options in the creation of authority records. Three options are: a) to create authority records only for complex headings, b) to create authority records for headings which require SEE or SEE ALSO references, or c) to create no authority records at all.

Subject headings reside in one physical authority file, and are identified in the bibliographic record by their indicator values and in the authority file by the byte in the fixed field. At Northwestern, the Medical and Transportation libraries have created separate subject authority records for their headings.

3.7.3 Links and references

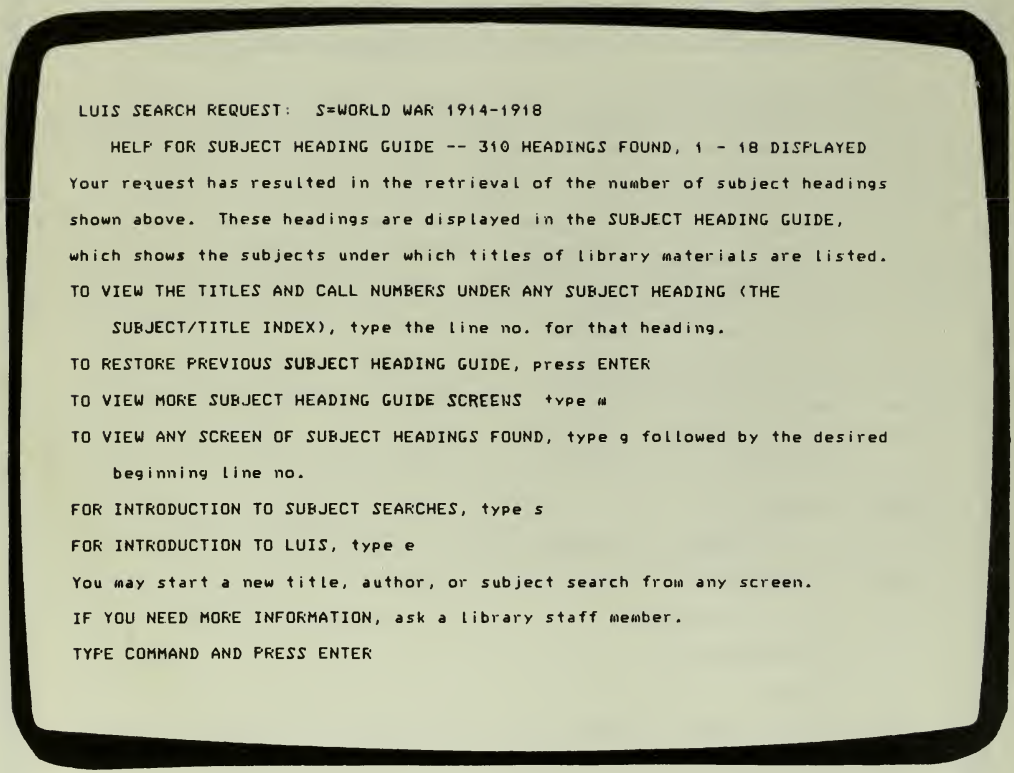
As authority records are created, the index entries for the 1xx, 4xx, and 5xx fields are incorporated into a NOTIS headings file. The headings still remain in the bibliographic record to protect the integrity of the file. NOTIS made an architectural decision to maintain headings both in bibliographic records and in their authority files. They are undergoing an index redesign, currently in test at Northwestern, however the same principle will apply. The relationship of the files will be specified by the clients. Most use a one-to-one (bibliographic to authority), while there are some many- (bibliographic) to-one (authority).

3.7.4 Retrieval and display

Subject searching is done only on the index for the subject headings in bibliographic records. Currently being tested at Northwestern is an index re-design that will allow users of the online catalog to search the indexes for both bibliographic and the authority files together, thus adding authority information such as see references and explanatory notes to the search results. At present, a partial subject match produces a browse of the index (referred to as a subject heading guide); an exact match displays brief titles and call numbers for records linked to the heading. These displays are shown in Figures N1-4.

At Northwestern, the Medical and Transportation Libraries do not use LCSH for subject analysis, but rely respectively on MeSH and a local scheme. To avoid confusion, subject headings for these two libraries are indexed separately and users are instructed to use special commands to access each index. That is, the command "s=" is used to search the LCSH index, "st=" retrieves transportation headings, and "sm=" accesses the index of MeSH terms. As shown in Figures N5-9, Northwestern has developed help screens to encourage users to learn more about the relevant thesauri.

```
LUIS SEARCH REQUEST   S=WORLD WAR 1914-1918
 SUBJECT HEADING GUIDE -- 310 HEADINGS FOUND, 1 - 18 DISPLAYED
  1 WORLD WAR 1914-1918
  2   --ADDRESSES ESSAYS LECTURES
  3   --ADDRESSES SERMONS ETC
  4   --AERIAL OPERATIONS
  5   --AERIAL OPERATIONS -BIBLIOGRAPHY
  6   --AERIAL OPERATIONS -CHRONOLOGY
  7   --AERIAL OPERATIONS BRITISH
  8   --AERIAL OPERATIONS CANADIAN
  9   --AERIAL OPERATIONS FRENCH
 10   --AERIAL OPERATIONS GERMAN
 11   --AERIAL OPERATIONS ITALIAN
 12   --AFRICA
 13   --AFRICA EAST
 14   --AFRICA FRENCH-SPEAKING WEST
 15   --AFRICA NORTH
 16   --AFRICA SOUTHERN
 17   --AFRO-AMERICANS
 18   --ALGERIA
TYPE m FOR MORE SUBJECT HEADINGS.  TYPE LINE NO. FOR TITLES UNDER A HEADING.
TYPE e TO START OVER.  TYPE h FOR HELP.
TYPE COMMAND AND PRESS ENTER
```


LUIS SEARCH REQUEST: S=WORLD WAR 1914-1918
HELP FOR SUBJECT HEADING GUIDE -- 310 HEADINGS FOUND, 1 - 18 DISPLAYED
Your request has resulted in the retrieval of the number of subject headings shown above. These headings are displayed in the SUBJECT HEADING GUIDE, which shows the subjects under which titles of library materials are listed.
TO VIEW THE TITLES AND CALL NUMBERS UNDER ANY SUBJECT HEADING (THE SUBJECT/TITLE INDEX), type the line no. for that heading.
TO RESTORE PREVIOUS SUBJECT HEADING GUIDE, press ENTER
TO VIEW MORE SUBJECT HEADING GUIDE SCREENS type m
TO VIEW ANY SCREEN OF SUBJECT HEADINGS FOUND, type g followed by the desired beginning line no.
FOR INTRODUCTION TO SUBJECT SEARCHES, type s
FOR INTRODUCTION TO LUIS, type e
You may start a new title, author, or subject search from any screen.
IF YOU NEED MORE INFORMATION, ask a library staff member.
TYPE COMMAND AND PRESS ENTER

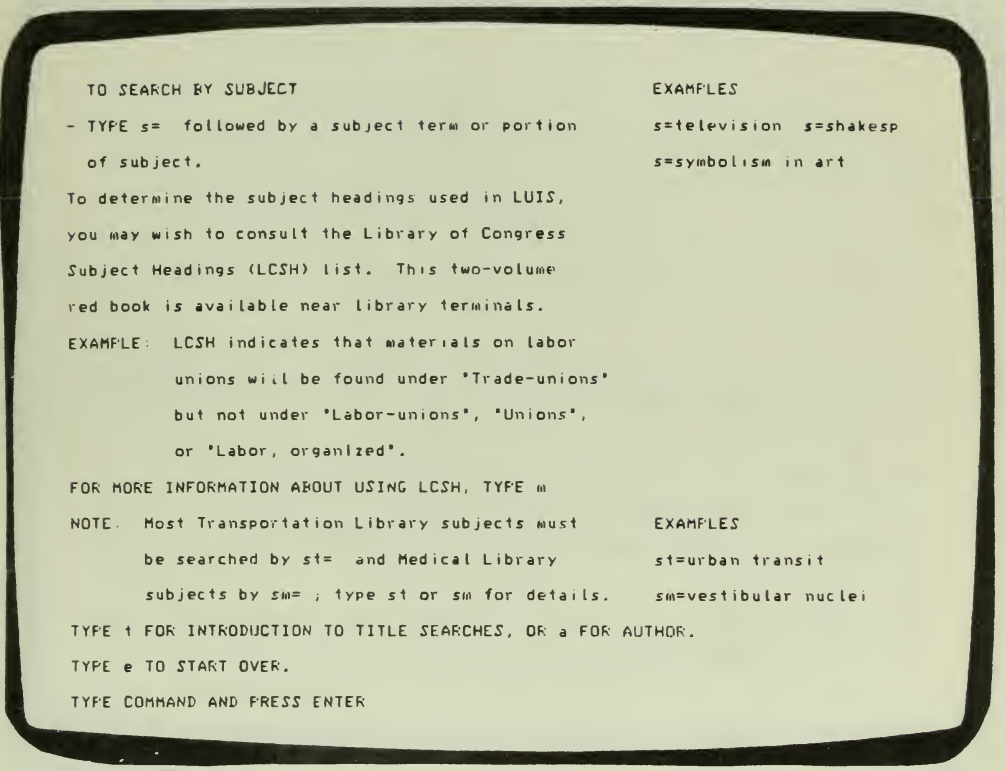
```
LUIS SEARCH REQUEST:  S=WORLD WAR 1914-1918
 SUBJECT/TITLE INDEX -- 80 TITLES FOUND, 1 - 16 DISPLAYED
WORLD WAR 1914-1918
  1 transactions of the grotius soci        NU main:341.06 G881
  2 first world war                  (1984)NU main:940.3 R634f
  3 world in the crucible 1914-1919  (1984)NU core:940.3 S355w
  4 world in the crucible 1914-1919  (1984)NU main:940.3 S355w
  5 churchills world crisis as histo (1983)NU main:940.45 C563Zp
  6 grande guerre                    (1983)NU main:940.3 M669g
  7 short history of world war i     (1981)NU main:940.3 S874s
  8 great war                        (1980)NU main:L940.3 B261g
  9 no mans land: 1918 the last year (1980)NU core:940.3 T647n
 10 no mans land: 1918 the last year (1980)NU main:940.3 T647n
 11 no mans land: 1918 the last year (1980)NU schf:940.3 T647n
 12 world war i: an illustrated hist (1980)NU main:L940.3 E93w
 13 zweite internationale und der kr (1979)NU main:324.1 B632z
 14 illustrated history of the world (1978)NU main:L940.4 I29
 15 italia unita e la prima guerra m (1978)NU main:945.09 R7631i
 16 lenin e mussolini: protagonisti  (1978)NU main:320.945 B793l
TYPE LINE NO. FOR BIBLIOGRAPHIC RECORD WITH HOLDINGS.
TYPE m FOR MORE TITLES.
TYPE g TO RETURN TO GUIDE.  TYPE e TO START OVER.  TYPE h FOR HELP.
TYPE COMMAND AND PRESS ENTER
```

```
LUIS SEARCH REQUEST:  S=WORLD WAR 1914-1918

    HELP FOR SUBJECT/TITLE INDEX -- 80 TITLES FOUND, 1 - 16 DISPLAYED

The subject heading you selected has listed under it the number of titles
shown above.  The entries in the SUBJECT/TITLE INDEX are displayed as follows:
- ORDER OF ENTRIES:  1.  periodicals and other serials (listed alphabetically),
  2.  books with publication dates (most recent listed first), 3.  books
  without dates (indicated by "n.d." and listed alphabetically).
- INFORMATION GIVEN:  title, publication date (for most books), code for library
  holding the item, location/call number.  Library codes:  NU (main & branches),
  GE (Garrett & Seabury-Western), LL (Law), TR (Transportation), HS (Medical).
    BIBLIOGRAPHIC RECORD OF TITLE WITH HOLDINGS, type line no.
TO RESTORE PREVIOUS SUBJECT/TITLE INDEX SCREEN, press ENTER
TO VIEW MORE SUBJECT/TITLE INDEX SCREENS, type m
TO VIEW ANY SCREEN OF TITLES, type i followed by desired beginning line no.
TO RETURN TO SUBJECT HEADING GUIDE, type g
FOR INTRODUCTION TO LUIS, type e ; TO SUBJECT SEARCHES, type s
You may start a new title, author, or subject search from any screen.
IF YOU NEED MORE INFORMATION, ask a library staff member.
TYPE COMMAND AND PRESS ENTER
```

```
LUIS SEARCH REQUEST:  S=WORLD WAR 1914-1918
BIBLIOGRAPHIC RECORD -- NO. 2 OF 80 ENTRIES FOUND
Robbins, Keith.
  The First World War / Keith Robbins. -- Oxford (Oxfordshire) ; New York :
Oxford University Press, 1984.
  186 p. ; 23 cm. -- (OPUS)
  Bibliography: p. (165)-171.
  Includes index.
 SUBJECT HEADINGS (Library of Congress; use s= ):
     World War, 1914-1918.
 LOCATION:  MAIN
 CALL NUMBER:  940.3 R634f
 CIRCULATION STATUS:  Not charged out. If not on shelf, ask at Circulation Desk
 LOCATION:  MAIN core  HAS BEEN ORDERED BUT NOT YET RECEIVED.
TYPE m FOR NEXT RECORD.
TYPE i TO RETURN TO INDEX, g TO RETURN TO GUIDE, e TO START OVER, h FOR HELP.
TYPE COMMAND AND PRESS ENTER
```


TO SEARCH BY SUBJECT
EXAMPLES
- TYPE s= followed by a subject term or portion of subject.
s=television s=shakesp
s=symbolism in art
To determine the subject headings used in LUIS, you may wish to consult the Library of Congress Subject Headings (LCSH) list. This two-volume red book is available near library terminals.
EXAMPLE: LCSH indicates that materials on labor unions will be found under 'Trade-unions' but not under 'Labor-unions', 'Unions', or 'Labor, organized'.
FOR MORE INFORMATION ABOUT USING LCSH, TYPE m
NOTE. Most Transportation Library subjects must be searched by st= and Medical Library subjects by sm= ; type st or sm for details.
EXAMPLES
st=urban transit
sm=vestibular nuclei
TYPE t FOR INTRODUCTION TO TITLE SEARCHES, OR a FOR AUTHOR.
TYPE e TO START OVER.
TYPE COMMAND AND PRESS ENTER

USING THE LIBRARY OF CONGRESS SUBJECT HEADINGS LIST

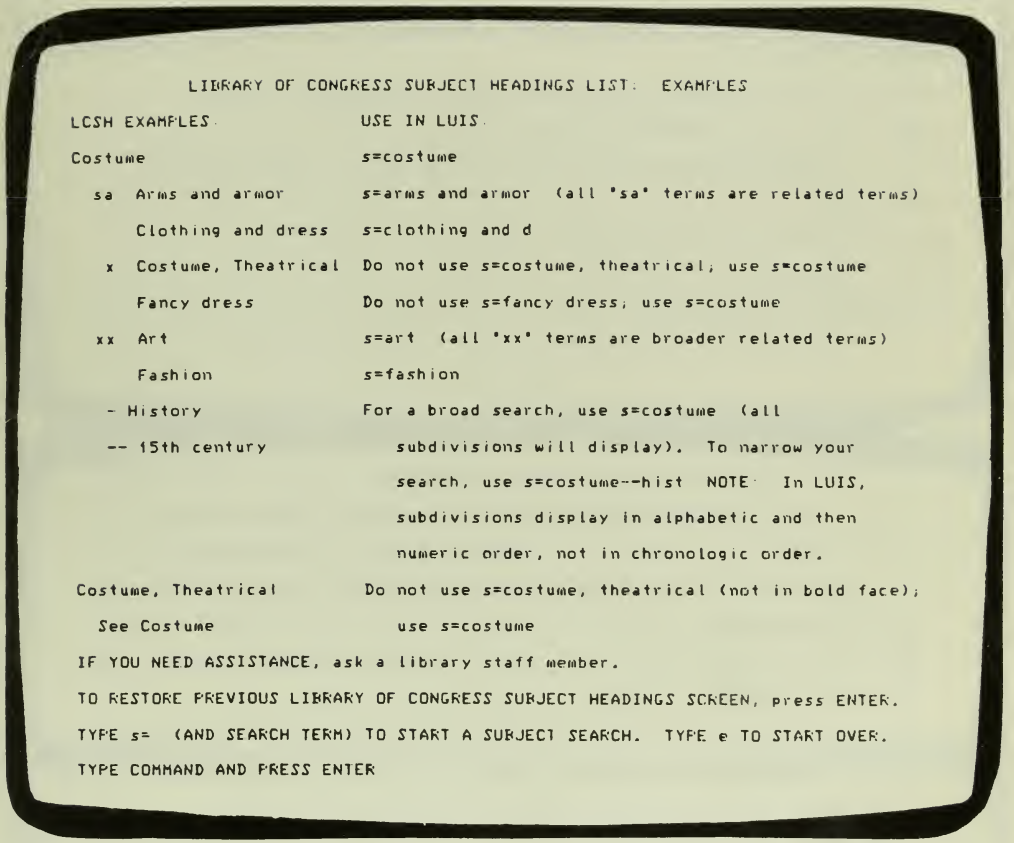
The Library of Congress Subject Headings (LCSH) list is an alphabetic guide to the forms of subject headings to be used when searching for materials by subject in LUIS or in the subject card catalog. When you consult LCSH you will find:

- Exact form of subject headings (in bold face type), which may be used to search in LUIS
- Subdivisions of those headings (preceded by a dash --), which will display automatically in LUIS under the term you enter if materials on those topics are in the database
- Related subject terms (preceded by sa for narrower or related terms or xx for broader terms), which may also be used to search in LUIS
- Terms which should NOT be used to search in LUIS (preceded by x)

TYPE m FOR LCSH EXAMPLES.

TYPE e TO START OVER.

TYPE COMMAND AND PRESS ENTER



LIBRARY OF CONGRESS SUBJECT HEADINGS LIST: EXAMPLES

LCSH EXAMPLES:	USE IN LUIS:
Costume	s=costume
sa Arms and armor	s=arms and armor (all "sa" terms are related terms)
Clothing and dress	s=clothing and d
x Costume, Theatrical	Do not use s=costume, theatrical; use s=costume
Fancy dress	Do not use s=fancy dress; use s=costume
xx Art	s=art (all "xx" terms are broader related terms)
Fashion	s=fashion
- History	For a broad search, use s=costume (all
-- 15th century	subdivisions will display). To narrow your search, use s=costume--hist NOTE: In LUIS, subdivisions display in alphabetic and then numeric order, not in chronologic order.
Costume, Theatrical	Do not use s=costume, theatrical (not in bold face);
See Costume	use s=costume

IF YOU NEED ASSISTANCE, ask a library staff member.

TO RESTORE PREVIOUS LIBRARY OF CONGRESS SUBJECT HEADINGS SCREEN, press ENTER.

TYPE s= (AND SEARCH TERM) TO START A SUBJECT SEARCH. TYPE e TO START OVER.

TYPE COMMAND AND PRESS ENTER

TO SEARCH BY MEDICAL SUBJECT HEADINGS

- Type sm= followed by a subject term or portion of subject.

EXAMPLES:

sm=neoplasms

sm=electrocardi

To determine the medical subject headings used in LUIS, you may wish to consult the Medical Library Subject Headings (MESH) list compiled by the National Library of Medicine. This book is available near terminals in the Medical Library and at service desks in the main and Science-Engineering libraries.

Most medical subjects must be searched by sm= . Some Medical Library items, however, have LCSH headings (use s=).

NOTE: MESH subject headings only are used in the Medical Library subject card catalog.

FOR MORE INFORMATION ABOUT USING MESH, TYPE m .

TYPE e TO START OVER.

TYPE COMMAND AND PRESS ENTER

USING THE MEDICAL SUBJECT HEADINGS LIST

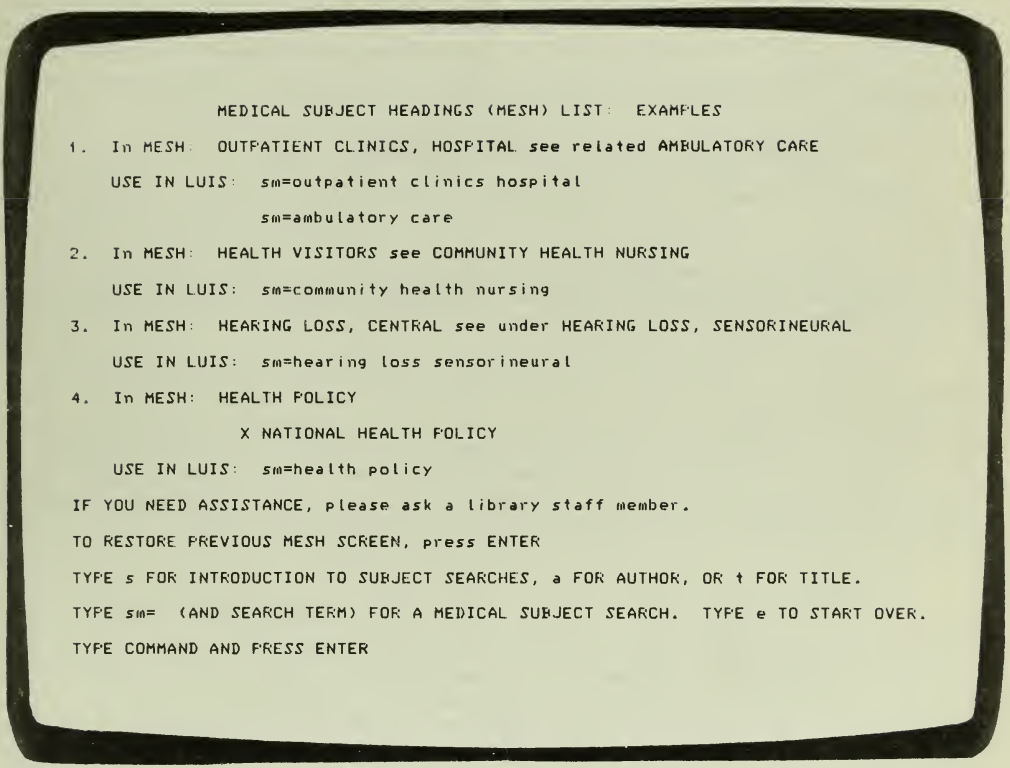
The Medical Subject Headings (MESH) list is an alphabetic guide to the forms of medical subject headings to be used when searching for materials by subject in LUIS or in the Medical Library subject card catalog. When you consult MESH you will find:

- Exact forms of subject headings (in bold face type), which may be used to search in LUIS
- Cross references which will guide you to subject headings which may also be used to search in LUIS:
  - see related refers to other subject headings which may be used
  - see refers from a subject heading which may not be used to one which may be used
  - see under refers from a subject heading which may not be used to a more general subject heading which may be used
  - headings preceded by X or XU may not be used
- The numbers in MESH cannot be used in LUIS.

TYPE m FOR MESH EXAMPLES.

TYPE e TO START OVER.

TYPE COMMAND AND PRESS ENTER



MEDICAL SUBJECT HEADINGS (MESH) LIST: EXAMPLES

1. In MESH: OUTPATIENT CLINICS, HOSPITAL see related AMBULATORY CARE

   USE IN LUIS: sm=outpatient clinics hospital

   sm=ambulatory care

2. In MESH: HEALTH VISITORS see COMMUNITY HEALTH NURSING

   USE IN LUIS: sm=community health nursing

3. In MESH: HEARING LOSS, CENTRAL see under HEARING LOSS, SENSORINEURAL

   USE IN LUIS: sm=hearing loss sensorineural

4. In MESH: HEALTH POLICY

   X NATIONAL HEALTH POLICY

   USE IN LUIS: sm=health policy

IF YOU NEED ASSISTANCE, please ask a library staff member.

TO RESTORE PREVIOUS MESH SCREEN, press ENTER

TYPE s FOR INTRODUCTION TO SUBJECT SEARCHES, a FOR AUTHOR, OR t FOR TITLE.

TYPE sm= (AND SEARCH TERM) FOR A MEDICAL SUBJECT SEARCH. TYPE e TO START OVER.

TYPE COMMAND AND PRESS ENTER

3.8 Carlyle Systems TOMUS

3.8.1 Background

TOMUS (The Online Multiple User System) is a turn-key library system designed and operated by Carlyle Systems, Inc. of Berkeley, California. The basic system is an online catalog and a modular design makes it possible for customers to buy only the portion of the system they need. There is a small version that runs on an IBM PC and larger versions that run on multiple processors. Carlyle builds the whole system, both hardware and software, and the system can be purchased, leased or rented. Their largest installation is CATNYP, The Online Catalog of the Research Libraries at New York Public Library.

3.8.2 Description of the subject authority file

TOMUS supports the MARC Format for Authorities. Authority records can be created in three ways: a) keyed directly into the system using the CARLYLE Cataloging System, b) obtained on tape from a commercial service, or c) automatically generated from the library's bibliographic records using the pre-processing module of the Carlyle Authority System.

Authority and bibliographic records exist as different formats within the same file. The system can support one or more logical subject authority files, but the different authority files will exist and be considered as one file for storage and maintenance. If there are identical terms in two authority files, there would be two authority records using the same term. Subject authority records from different source lists are distinguished by byte 11 in the fixed field.

Theoretically there are no limits to the number of files that can be maintained; the only limit is the desire of the client to maintain them. One of Carlyle's newest clients, St. John's University in Minnesota, has been entering Catholic subject headings as 690's in their edited version of OCLC records. Their application of TOMUS will test the capability to handle multiple subject authorities as records will need to be retrieved using either Catholic (Kapsner) or LC subject headings.

3.8.3 Links and references

Libraries define authorized headings as those they wish displayed, based on tags and indicator values (i.e., 650 or 690, indicator 0, 1, 2, etc.). Headings are maintained in the bibliographic record rather than through a link to an authority file, and cross-references are added to bibliographic records when appropriate. As records are loaded, codes preceding a field tag identify four different situations:

1) a heading in a bibliographic record matches an authorized form in the authority file (identified in the bibliographic record as e.g., a110)

2) a heading in a bibliographic record matches a cross-reference in the authority file and the authorized form of the heading is copied from the master authority record to a specially-created field in the bibliographic record (again identified as e.g., a110)

3) a heading in a bibliographic record may match no heading in the authority file (identified in the bibliographic file as a provisional heading and marked as e.g., p110); bibliographic records with provisional headings are copied to a separate file for review by the library

4) a heading in a bibliographic record matches two or more headings (authorized or cross-referenced) in the authority file (identified as a conflict and marked as e.g., c110); bibliographic records with conflicting headings and their associated authority records are copied to a separate file for review by the library.

Since the authorized headings are maintained in the bibliographic record, it is possible to authorize the records in a particular bibliographic file against more than one subject authority file. Different headings in the same record can also be matched to different subject authority files.

Authority records cannot be linked to each other. There are no edit checks or maintenance capabilities for RT, BT, NT references added to records for those headings, however keyword searching makes it possible to find subdivisions of headings. Since the authorities are supported in one file, there are no links made across different authority files.

3.8.4 Retrieval and display

The authority file index is automatically searched together with the bibliographic file indexes. There is no browsing of separate subject authority files or indexes. If an authorized heading matches the search, the bibliographic records are automatically displayed. If a search matches a see reference, the authorized form of the heading is automatically searched and the associated bibliographic records are retrieved. The searcher is notified about retrievals on both terms. An example of a display that would direct users from non-authorized forms of a subject to authorized forms follows:

```
Items on the subject
     HEART ATTACK
are cataloged by this library under the term
     CORONARY HEART DISEASE
To search for these items, press the key marked RETURN.
```

All fields in a bibliographic record can be searched in the TOMUS system. The library can define the various indexes to be searched and specify the displays. If properly tagged, subject

headings from different thesauri could be defined in separate indexes (e.g., author, title, LC subject, medical subject) and searched using separate commands. A library also could define what headings are considered authorized. In the "full" and "brief" displays, only authorized headings will display. In the MARC display, all headings would appear. In the case of St. John's University, the library will be determining the selection of the authorized displays. For example, a search on "Lord's Supper" will result in an authorized heading of "Lord's Supper" in LCSH, but the words are a cross-reference to "Eucharist" in the Kapsner file. Since both terms have been used and designated as authorized, some means of specifying the cross-reference relationship between the two subject authority files will need to be made. One possibility was to create separate indexes and have the system search one subject file first. This idea was discarded because it would not provide a means to ensure that searchers would continue on with a search, nor could they be automatically referred to the correct term in the other index.

The key ingredient to the TOMUS subject search capability is that the system is based on keyword searching. Users may search on any significant word and searches may be limited by date, language, country of publication, format or medium. In addition, users may combine searches using Boolean logic. TOMUS's designers note that the use of an authority file to control subject vocabulary must be adjusted in the context of keyword searches. The keyword search capability makes it possible for users to retrieve records without needing to know the correct order of terms used as the authorized heading and the system need not provide see references from inverted terms. When word order is the only difference between terms from different thesauri (e.g., "Clinical psychology" and "Psychology, Clinical"), no link between them is needed.

Chapter Four
Subject Searching

4.1 Online catalog search features

In many respects, subject searching in online catalogs always involves manipulation of multiple vocabularies--the users' vocabularies and the library's controlled vocabulary. Retrieval design features significantly affect a system's ability to match user inquiries to controlled subject terms. The extent to which multiple controlled vocabularies can be used effectively in a particular system will be governed in part by the basic searching features of that system. Relevant features include: 1) index construction, 2) displays, 3) subject indexing browse, and 4) use of a subject authority file.

Indexing, or the construction of subject search keys, is one of the most powerful determinants of the success of a subject search. Few online systems restrict searching to the exact left-to-right, word-for-word matching required by manual files, since machine searching must compensate for the loss of discriminating browsing possible in eye-readable catalogs. On the other hand, large indexes are relatively expensive to create, maintain, and search; thus, many existing systems lack full key word searching of subject headings. Figure 4-1 describes some of the subject search matches available in online catalogs; the options listed are not mutually exclusive. Matches that are less precise (i.e., more forgiving) are more likely to bring up a variety of terms rather than resulting in one or no hits. This greater flexibility can make a mix of vocabularies more tolerable by eliminating some of the distinctions between terms from different lists. For example, an exact match search of the MeSH term "Psychology, Clinical" would miss records assigned the LCSH term "Clinical psychology." A component word search, on the other hand, would ignore the variation in word order between the two lists and retrieve both terms.

Screen displays of search results vary widely in layout, data elements shown, language used, explanations provided, etc. The design of displays (also affected by what data elements are available for display) can have a major impact on a user's ability to interpret terms retrieved from multiple subject sources. Figure 4-2 illustrates displays for a search that retrieved 10 records using the MeSH term "Psychology, Clinical" and 2 records with the LCSH term "Clinical, Psychology." These displays are based on ones used in actual systems. The examples illustrate a considerable range in the degree of assistance they provide for the user.

The user's ability to select subject terms and interpret search results is enhanced by the availability of a browsable subject index. In some systems, an index display is available only when a user's search matches more than one term; in other systems, certain commands will always bring up an index array of

subject terms. Index displays may also include the number of postings in the database to aid users in selecting terms. The nature of the index display varies considerably, depending whether the search key has been a strict left to right match or a match to a word or subdivision within a heading. In the former case (left to right match), the system will usually display an alphabetical sequence of terms that includes the user's search. In the latter case (component word or subdivision match), the display will be a KWIC listing of heading that contain the user's term. Examples are shown in Figure 4-3. Some systems provide both displays in response to different commands. Index displays can include terms from more than one subject list. The terms may be interfiled (with or without indication of source list) or displayed in separate sequence. Examples of separate sequences are shown in the WLN figures W-4, W-6, and W-7 in Chapter Three.

The fourth major feature affecting subject searching is the availability of a subject authority file. The authority file has the potential to provide two significant retrieval aids: cross-references and displays of related terms. Cross-references are displayed in a variety of ways, and may provide an automatic retrieval on the authorized term. Cross-references are often, but not always, included in index displays. Related terms are usually displayed in response to specific commands, such as the "expand" and "expandall" commands described in the WLN section of Chapter Three. In another example shown in Figure 4-4, users of the Ohio State University online catalog are given the option to view related terms and the subject authority information using the "SAL" command. Access to a subject authority file can both enhance and complicate searching when multiple thesauri are included in the database. Cross-references and related term displays may appear to conflict unless the subject source for each term or reference is clearly indicated and the authority file for each list is logically separate. The systems described in Chapter Three provide searching via multiple subject authority files. However, none has yet developed displays and help screens that can extend this sophisticated search feature to a typical end-user.

4.2 Approaches to Retrieval of Multiple Subject Vocabularies

Essentially, there are four basic approaches for providing end-user access to databases indexed by different controlled vocabularies. Each of these poses different possibilities and design problems for online library catalogs. These approaches can be summarized as: 1) segregated files, 2) mixed vocabularies, 3) integrated vocabularies, 4) front-end navigation.

Segregated files. In this approach, collections using different subject thesauri are defined as separate catalogs, at least for the purposes of subject searching. The system would maintain logically separate subject indexes and subject authority files for each vocabulary. This approach works best when the different vocabularies are associated with distinctly different files (akin

to separate library catalogs) and when users are most often interested in specific files. For example, Northwestern University's online catalog, LUIS, instructs library patrons to use different commands when searching subject headings for the main library, the medical library, or the transportation library. Most existing online catalogs do not support segregated subject searching within an otherwise unified database. However, in cases where variant thesauri are restricted to use with specific formats (e.g., visual materials) and/or specific locations (e.g., a branch library), searches limited by format or location could provide the needed definition.

The segregated approach enables searchers, particularly those who are experienced, to take full advantage of well-developed specialized vocabularies that have been applied to special collections. The obvious disadvantage is the requirement for multiple sequential searches when users are interested in materials in more than one collection.

Mixed vocabularies. The extent to which unrelated controlled vocabularies can productively cohabit the same online catalog is dependent on a large and complex mix of variables. These variables include the many retrieval design options described in the first part of this chapter. They also include the skills and needs of users, and the number and nature of the actual vocabularies in question. In general, two kinds of problems must be overcome for the mix to be successful: 1) vocabulary clashes, and 2) degradation of access to specialized collections.

Vocabulary clashes are inevitable when terms from two unrelated thesauri are mixed. The term "clash" is significant here, intended to focus on retrieval results that are conflicting or distinctly misleading. The assumption is that more subtle differences in thesauri, such as varying levels of specificity or variant subdivision practice, can be tolerated in a library catalog where changing practices, limited headings maintenance, and variations in catalogers' judgments have already created a somewhat imprecise tool. It can even be argued that the enriched vocabulary from additional sources compensates for the loss of precision in application. However, particular kinds of obvious vocabulary clashes will cause problems in retrieval. Clashes include:

1. Conflicting cross-references, i.e., a postable term in one thesaurus is a non-postable term in another.

2. Homographs, e.g., "Irrigation" as it is used differently in MeSH and LCSH.

3. Different related term structures for the same term used in both thesauri.

The degree to which this apparently conflicting information can be tolerated in an online catalog depends on the mix of design features supported by the retrieval system. Figure 4-5

uses a simplified hypothetical example to illustrate how design choices can affect the user's ability to interpret clashing results. At minimum, users should be able to browse subject indexes that contain both headings and cross-references and to be able to distinguish easily among results from different lists. The situation is easiest to cope with when different subject terms can be associated with different collections.

The other serious problem posed by mixing vocabularies in online catalogs is the potential for impeding access to smaller, specialized collections. The most common mix of thesauri in library catalogs includes the headings used by the main collection (typically LCSH) and a limited set of thesauri (often just one other) for smaller special collections, such as a children's room or medical library. Presumably the specialized vocabulary has been applied because it better serves the clientele of the special collection. In cases where users are interested only in using the smaller collection (e.g., they are interested only in graphic materials), the inclusion of terms from other lists will, at best, slow down and complicate subject searching. At worst, users can be misled when a search produces a hit in the main library and the term appropriate for use in the specialized collection is missed. In cases where the union catalog must be used as the access tool for a specialized collection, it is important to provide the option to limit subject searching by vocabulary and/or collection.

Integrated vocabularies. Chapter Two describes five techniques for relating different thesauri. These techniques could be used to develop references that would link related terms in different thesauri and would aid retrieval in those online catalogs that employ an authority file in subject searching. Intellectual mapping of one vocabulary to another, or to a switching language, is an arduous task. On the other hand, it is possible to envision the practical application of a limited form of vocabulary integration, one that would at least catch vocabulary clashes. Vocabulary integration involves the creation of a master file of terms and references in the thesauri in question. Machine matching can identify similar/identical terms (including homographs) as well as postable and non-postable terms that conflict. A combination of computer processing and editorial review can be used to supply "see also" references between otherwise conflicting terms. The National Library of Medicine has already performed this level of integration between MeSH and LCSH. ORION designers plan to use NLM's work to create "Search also under" links between MeSH and LCSH terms that would otherwise appear to conflict in UCLA's online catalog. Since ORION subject searches provide the user with an index listing of terms and references, these links will aid users in selecting terms.

In online catalogs supported by subject authority files, headings requiring this limited integration could be automatically identified when new authority records were added or loaded. Provisional related term references could also be

generated automatically. However, this is a feature most practically operated as part of the thesaurus/authority file maintenance function, since it cannot be completed without some editorial review. Chapter Two enumerates system features that support thesaurus integration, particularly the limited integration that would deflect vocabulary clashes in online catalogs.

Front-end navigation. The last decade of research has produced a number of experimental or prototype systems designed to match users' search terms to terms in one or more controlled vocabularies. An experimental system created at the Massachusetts Institute of Technology, CONIT, received considerable attention for its exploration in mapping key words and word stems from users' search terms to the controlled vocabularies in multiple online databases. The experimental Vocabulary Switching System (VSS), developed by Niehoff at Batelle, maps or switches users' terms to terms in 15 controlled vocabularies. (10) VSS offers 21 options for "vocabulary switching," ranging from exact matching on lead terms, to stem matches, to exhaustive matches on all possible synonyms and related terms. This complex system, designed to test and compare various switching options, is not in its current form a prototype end-user interface. A successful operational system is Paperchase, which is focused on a very limited application. Developed to assist physicians at Beth Israel Hospital in Boston in searching a subset of the MEDLINE database, Paperchase normalizes users' search terms and attempts to match them to MeSH terms and rotated versions of MeSH terms.

Term matching is not the only kind of assistance that can be provided by an "intelligent" interface. Techniques such as weighting terms by frequency of use in the database or automatically seeking documents similar to those judged relevant by the user can also aid retrieval, as was demonstrated by NLM's prototype online catalog, CITE. These techniques begin to move into the realm of designing expert systems for searching, systems that can navigate the user through the kinds of search strategies that an expert intermediary would employ. Powerful microcomputers can serve as intelligent terminals in which interface software can reside, enabling the design of interfaces that need not affect the programming of the retrieval system itself. Even on a less "expert" level, microcomputer interfaces can contain menus and help screens to aid users in displaying and selecting terms from multiple online thesauri. Such front-end navigation holds promise for searching multiple vocabularies in future online catalogs.

Figure 4-1

Examples of Subject Indexing in Online Catalogs

1. Normalization of punctuation and capitalization (e.g., a search on "BIRD-HOUSES" retrieves "BIRD HOUSES").

2. Left-to right match of entire subject string (e.g., must know entire correct LCSH term including subdivisions in correct order).

3. Left match of main term, ignoring subdivisions (e.g., search on "CATS" also retrieves "CATS-BIBLIOGRAPHY").

4. Match of main term and some/any subdivisions (e.g., search on "CATS-BIBLIOGRAPHY" also retrieves "CATS-U.S.-BIBLIOGRAPHY").

5. Left match of lead term(s) with implicit or explicit truncation (e.g., search on "CATS" also retrieves "CAT").

6. Exact match on main term or subdivision (e.g., search on "BIBLIOGRAPHY" retrieves "BIBLIOGRAPHY-U.S." AND "CATS-BIBLIOGRAPHY").

7. Matches on component word combinations (e.g., search on "DOGS AND CATS" also retrieves "CATS AND DOGS").

8. Matches on any component word (e.g., search on "DOGS" also retrieves "CATS AND DOGS").

Figure 4-2

a) Clinical psychology (2)

Psychology, Clinical (10)

b) Clinical psychology (2)

M Psychology, Clinical (10)

c) Your search: FIND SUBJECT CLINICAL PSYCHOLOGY
Found items under the following terms

Clinical psychology (Main library subject)	2
Psychology, Clinical (Medical library subject)	10
Total items found	12

Figure 4-3

a. Alphabetical index display of search on "Psychology"

Psychobiology
Psychodrama
Psychology
Psychology--Congress
Psychology, Clinical
Psychotherapy

b. KWIC index display of search on "Psychology"

Art--Psychology
Clinical psychology
Librarians--Psychology
Psychology
Psychology--Congress
Twins--Psychology

The Ohio State University LIBRARIES

LCS (Library Control System)

SUBJECT HEADINGS INDEX DISPLAY (PUBLIC)

```
 01        Library duplicates
 02           SEARCH UNDER:  Duplicates in libraries
 03        Library economy
 04           SEARCH UNDER:  Library science
>05     48 Library education        *(SEE BELOW)
 06      6 Library education--Addresses, essays, lectures
 07      1 Library education--ASIA
 08      1 Library education--AUDIO-VISUAL AIDS
 09      1 Library education--AUDIO-VISUAL AIDS--BIBLIOGRAPHY
 10      1 Library education--Brazil.
ENTER TBL/line no. FOR TITLES.  *ENTER SAL/line no. FOR MORE INFORMATION
ENTER PS- FOR PRECEDING PAGE; ENTER PS+ FOR NEXT PAGE.
sub/library education
```

RELATED HEADINGS AND NOTES (PUBLIC)

```
Library education
   POSSIBLE BROWSING NUMBER(S):  Z668-9  (SEARCH WITH SPS/)
   Here are entered works on the education of librarians. Works dealing with
      the instruction of readers in library use are entered under the heading
      Libraries and readers.
   SEARCH ALSO UNDER:
      Interns (Library science) (2 TITLES)
      Librarians--In-service training (10 TITLES)
      Library employees--In-service training (1 TITLE)
      Library orientation (29 TITLES)
PG1.  ENTER PG2 FOR NEXT PAGE.
ENTER PS1 TO RETURN TO LIST OF SUBJECTS.
sal/5
```

SUBJECT HEADINGS INDEX DISPLAY (STAFF)

```
01          1867108    Library duplicatee
02           258763       SEARCH UNDER:  Duplicates in libraries
03          1880053    Library economy
04            30654       SEARCH UNDER:  Library science
>05    48     11555    Library education       *(SEE BELOW)
06     6     183328    Library education--Addresses, essays, lectures
07     1      85602  * Library education--ASIA
08     1      37820  * Library education--AUDIO-VISUAL AIDS
09     1     170401  * Library education--AUDIO-VISUAL AIDS--BIBLIOGRAPHY
10     1    2256624  * Library education--Brazil.
ENTER TBL/line no. FOR TITLES.  *ENTER SAL/line no. FOR MORE INFORMATION.
ENTER PS- FOR PRECEDING PAGE; ENTER PS+ FOR NEXT PAGE.
sub/library education/all
```

RELATED HEADINGS AND NOTES (STAFF)

```
Library education   (GEOG)  (1981)
02    11555   SUB: 48
03 (053) Z668-9
04 (680) Here are entered works on the education of librarians. Works dealing
           with the instruction of readers in library use are entered under
           the heading Libraries and readers.
07 SEE FROM: 5
08  1880017  Education for librarianship
09  1880019  Librarians, Education of
10  1880018  Librarians, Training of
PG1.  ENTER PG2 FOR NEXT PAGE.
```

```
sal/5/all

Library education        11555
12 SEE FROM:
13  1880020  Library school education
14  1880021  Library science--Study and teaching
15 SEE ALSO: 5
16   220192  Interns (Library science) (2)
17    54673  Librarians--In-service training (10)
18  1352822  Library employees--In-service training (1)
19    48231  Library orientation (29)
20    11554  Library schools (3)
PG2.  ENTER PG1 FOR PRECEDING PAGE; ENTER PG3 FOR NEXT PAGE.

pg2
```

Figure 4-5

Hypothetical search on "CATS" where "CATS" is the correct term in the main library, "CAT" is used in the Veterinary Library, and "Felis catus" used in the Zoology Library.

a. Systems displays bibliographic hit only

CATS 1
(User misses 6 records in Veterinary Library and 3 in Zoology)

b. System displays hits on bibliographic and authority records:

CATS 1
CATS Search under Felis catus
(User sees unexplained conflicting information and misses records in Veterinary)

c. System displays a subject index:

CARS	12
CAT	6
CATS	1
CATS Search under FELIS CATUS	3

(User sees that there are hits on a number of related terms)

d. System displays a subject index and distinguishes among vocabularies:

CARS (Main)	12
CAT (Vet)	6
CATS (Main)	1
CATS Search under FELIS CATUS (Zoo)	3

(User sees that there are hits on a number of related terms; trained users will understand why terms vary.)

Chapter Five
Support for Multiple Thesauri at the Library of Congress

5.1 Current Situation

Like other large research libraries, the Library of Congress has developed a number of special collections that are accessed by specialized subject vocabularies. Unlike other libraries, LC is responsible for the maintenance and publication of what is probably the world's largest controlled vocabulary, LCSH. In addition, the Library produces two printed bibliographic publications--the National Union Catalog of Manuscript Collections (NUCMC) and the Handbook of Latin American Studies (HLAS)--that maintain their own terminology for subject indexing. The nature and scale of LC's controlled vocabularies vary widely, as do their methods of maintenance and application. Each is described in the following sections.

5.1.1 Library of Congress Subject Headings

The Library of Congress Subject Headings (LCSH) is by far the largest, most comprehensive, and most widely used thesaurus maintained by the Library of Congress. The list is used as the source for subject headings applied to materials in all formats and all subject areas cataloged by LC's Processing Services (with the exception of the National Union Catalog of Manuscript Collections) and included in MARC records distributed by the Cataloging Distribution Service. The list supplies subject headings for records in thousands of North American library catalogs. Some 145,000 terms are included in the 9th edition of LCSH, with about 5,000 added each year. Many more unique, authorized LC subject headings exist on MARC records--more than two million are in the OCLC database of member and LC MARC records--since not all possible names or combinations of headings and subdivisions are specified as terms in LCSH.

Creation and maintenance of LCSH is based in LC's Subject Cataloging Division, where all subject catalogers contribute to the creation of new terms and references and a staff of six is responsible for editing their proposals. From 1898 to 1985, LCSH was maintained in card form. The published LCSH was produced separately, from GPO print tapes during the 1960's and since 1973 from tapes input and maintained through batch data processing at LC. In 1986, the master file of print tapes was converted into a record structure compatible with the U.S. MARC Authorities Format and loaded into LC's MUMS system, allowing LCSH records to be searched, input, and updated online. With the implementation of Subjects Release 1.0, MUMS now provides basic thesaurus maintenance support for LCSH. Output from the MUMS Subject Authority File will be used to produce the print and microform publications of LCSH and to create tapes in the US MARC Authorities format for distribution by the Cataloging

Distribution Service. At LC, subject authority records are searchable online by anyone using MUMS; subject authority records may be loaded into SCORPIO at a later date.

Compiled since 1898, LCSH has not been designed to conform to currently accepted standards for thesaurus construction (ANSI Z39.19). While there have been calls from library professionals for major revisions of LCSH to bring it up to current standards, the implications of such a revision are great--both in terms of the cost of such a project and, more important, the effect such change would have on existing library catalogs. Instead, more modest, incremental efforts have been made to align the list with current standards as new terms are added, including an increasing use of natural word order and more careful articulation of broader/narrower/related term relationships. In addition, when the master tape file of LCSH records was converted for loading into MUMS, coding in the 5xx control $w byte "0" was set to indicate whether "see also from" references were "broader" terms or "related" terms, a distinction recommended in the ANSI standard. References to narrower terms (i.e., 3xx fields) were not carried into the converted file; these may be assumed to be reciprocals of the appropriately coded 5xx fields and as such can be machine-generated for displays and other system requirements.

5.1.2 Legislative Indexing Vocabulary

The Legislative Indexing Vocabulary (LIV) provides controlled subject descriptors for the online reference files and subject catalogs created and maintained by the Congressional Research Service, including bibliographic files (citations to journal literature, documents, reports) and text files such as bill digests. The thesaurus is accessible online in SCORPIO (using the LIVT or EXPN command) and in a printed version cumulated and published annually. It is also maintained in an online LEXICO file. The thesaurus is compiled and maintained by an editor based in the Library Services Division of the Congressional Research Service. It is used by about 40 staff members who index material for CRS databases.

The thesaurus consists of some 9,270 terms, of which 54% are postable and 46% are non-postable (i.e., are "USE" references). According to the published introduction, "the scope of LIV encompasses the broad research areas of public affairs assigned to the research divisions of the Congressional Research Service, stressing the social sciences as well as policy aspects of the pure and applied sciences." Terms selected must have literary warrant in materials covered by the CRS databases and are formulated with the needs of a specific user group in mind. The list is relatively new by LCSH standards, with the first edition published in 1970. Terms and references have been carefully edited in conformance with standard contemporary thesaurus construction practices (i.e., ANSI Z39.19), including direct entry word order and use of BT, NT, RT and USE/UF references.

At present, LIV is maintained in two separate, parallel files. The thesaurus is created and edited in a batch production system to create tapes for loading into SCORPIO and for issuing a print product. The batch system also provides statistics and special listings such as a monthly list of postable terms. LIV is also maintained in a separate LEXICO file. LEXICO provides the editor with many desirable maintenance features, such as online access and browsing of the file (SCORPIO searches and displays only one LIV term at a time), automatic checking of reciprocal terms, etc. Unfortunately, no appropriate machine-readable output or interface is yet available from LEXICO for use at LC. Neither LEXICO nor the LIV batch system accepts, maintains, or outputs term records in the MARC Authorities Format.

Compatibility with LCSH is an expressed consideration in selecting terms for LIV. LCSH is searched before each LIV term is established and about 70% of LIV terms are compatible with LCSH main terms. Each LIV term is assigned an LCSH compatibility code as follows:

LC - term is the same in LCSH;
LCX - term is a "see" reference in LCSH;
LCC - term is similar to one in LCSH;
LCD - characters match an LCSH term, but the meaning is different;
LCO - no match in LCSH.

Typically variations from LCSH result from: LIV's adherence to direct word order, the appearance of a concept in the CRS-indexed periodical literature and reports preceding use in LCSH, and differences in LIV subdivision practice.

5.1.3 Thesaurus of Graphic Materials

The Library of Congress Thesaurus of Graphic Materials : Topical Terms for Subject Access (TGM) provides over 5,200 controlled subject descriptors (counting both postable and non-postable terms) for the cataloging records created by the Prints and Photographs Division of Research Services. The Division creates some 250 records per year describing groups of pictorial material and approximately 3600 records per year for individual items within the collections. Since February 1986, P&P's group-level records have been created in the MARC Format for Visual Materials, added to LC's MUMS file, and will be distributed on tape by the Cataloging Distribution Service. The records carry TGM terms in the 650 field with the second indicator coded as "7" and "lctgm" specified in subfield 2. Records for individual items are currently created for only P&P's card catalog, using TGM terms as subject headings. In addition, items reproduced on the Division's videodisks are being indexed using BRS software on a microcomputer and TGM terms are used as descriptors in this database.

The TGM is created and maintained in the Prints and Photographs Division by picture cataloging staff, which does both descriptive and subject cataloging of the materials in the Division's custody. The thesaurus is maintained online using the LEXICO system which provides critical editorial support features such as online editing and input, automatic checking and insertion of reciprocals, etc. However, since LEXICO has no machine-readable output or interface capability, the thesaurus cannot be made available for searching via MUMS or SCORPIO. LEXICO currently permits limited profiling for printed output of term lists. This output is being used as the master for a published TGM list slated for distribution by CDS in the fall of 1986. The Division receives frequent (approximately five per month) requests for copies of TGM from catalogers of visual materials collections, who await the published list.

Started in 1983, TGM evolved from the 1981 List of Subject Headings used in the Library of Congress Prints and Photographs Division. In its present version, TGM is formulated according to the current ANSI standard for thesaurus construction, including terms in natural word order and specification of term relationships into RT, NT, BT, and USE/UF. A small number of standard subdivisions are; otherwise subdivision is only by chronological and geographic specification.

Both LCSH and LIV are routinely checked before candidate terms are approved for inclusion in TGM, and most TGM terms are derived from LCSH. An LCSH term may be altered to conform to the ANSI rules of syntax, or to incorporate more up-to-date terminology. When this is done, a USE reference is usually established from the LCSH term to the TGM form. When the TGM LEXICO system is fully operational, an LCSH compatibility code will be added for each TGM term record in a manner similar to LIV's LCSH compatibility coding.

Despite efforts to maintain compatibility, there are significant differences between TGM, a list created to describe the content of pictures, and LCSH, a list designed to convey the content of books. First, many terms in LCSH are too specific and overlap in definition and scope in relation to images being cataloged. Second, many LCSH terms related to pictorial material are inappropriate for graphic materials. For example, the subdivision "Pictorial works" or the modifier "Pictorial" would apply to virtually all material cataloged; the free-floating phrase "in art" implies a value judgement not necessarily appropriate for documentary graphic materials. Finally, LCSH lacks terms for concepts that are primarily visual in nature and those that are too specific to be the subject of a book but are frequently subjects sought in pictures. Examples are: "Mushroom clouds", "Hammer and sickle", "Ship of State", and "Moonlight."

5.1.4 Subject Headings for Children's Literature

The Subject Headings for Children's Literature is a supplement to LCSH containing 400-500 terms that extend or differ

from terms authorized by LCSH. The supplement is maintained by the six-member Children's Literature Section of the Subject Cataloging Division, which uses LCSH and the supplement as the source for subject terms assigned to materials included in the Annotated Card (AC) Program.

Approximately 3,000 titles per year are selected from current LC receipts for AC handling; these are materials considered likely to be added to school libraries. Records for AC materials are amplified by selected added entries, an annotation describing the work, and an additional set of subject headings. These records are included in LC's MARC database, distinguished by the suffix "AC" added to the LC card number. AC subject headings are identified by the use of the second indicator "1" in the 6xx field of MARC bibliographic records. When the AC program uses an LCSH term, it repeats the term in the AC "set," coding it with the second indicator of "1." The LC MARC database currently contains approximately 60,000 records with supplementary AC data.

AC headings were developed in 1965 to provide easier, more appropriate access to materials used by children. The first published list appeared in 1969, and since then has been considered a standard list for use in cataloging children's materials. AC headings are maintained in a card file which contains both bibliographic and authority records for all AC titles. Currently this card file represents the only separate file of headings used by the AC program, since AC records are not segregated for searching in LC's online file.

The AC printed supplement had been produced by the LCSH batch system. With the implementation of LCSH Online (which excludes the AC exceptions) the Children's List will require another means for preparing the published list. Since the list is small, editing and production can be supported on a variety of word processing or microcomputer systems. However, because the Children's List is supplementary to LCSH, intellectual maintenance of the list requires access to LCSH and to a file of those LCSH terms used by the AC program, as well as to the supplementary AC list itself.

A "see" reference is made from an LSCH term whenever an alternative form is selected for the Children's List (with the exception of some differences in punctuation only). Headings recommended for the AC list are processed through the same editorial reviews as LCSH terms. Proposed AC terms are often adopted by LCSH, keeping the exceptions to a minimum.

Differences between LCSH and AC headings are spelled out in the introduction to the LCSH 10th edition. In summary, these differences include: modernized spelling, preference for subdivisions rather than inverted forms ("Animals--Training", instead of "Animals, Training of"), removal of qualifying terms such as "Children's" or "for children," use of common rather than scientific names, adoption of some Sears terms, introduction of

terms for concepts lacking in LCSH, and variations in subdivision practice. Other variations are in application rather than form, such as the obvious elimination of the subdivision "juvenile literature," and the infrequent use of "American" or "United States" since most U.S. children's works reflect an American orientation. Differing application policies mean that identical headings on AC and other LC MARC records may not reflect the same subject coverage when these headings are retrieved together in a bibliographic file.

5.1.5 National Union Catalog of Manuscript Collections

The National Union Catalog of Manuscript Collections (NUCMC) is a printed catalog of entries describing groups or collections of manuscripts held in repositories throughout the United States. Annual listings are accompanied by an index volume which provides in one alphabet references to names, places, and subjects reported in NUCMC entries. The indexes cumulate over a five year period.

Since the index volumes integrate all forms of access points, no separate subject headings list file or thesaurus is maintained by NUCMC editors. An authority card file is maintained for all types of headings to show cross-reference usage and verification research as needed. Authority control is provided by editing headings destined for the new index volume against three files: the authority file, the interfiled entry cards that will be used to print the next cumulated index, and a printed volume. As current manual practices are changing (drastically) with the implementation of RLIN cataloging, NUCMC staff is in the process of establishing new authority files. The NUCMC index is maintained by the six-member Manuscripts Section of the Special Materials Cataloging Division of Processing Services.

Since NUCMC subjects are not maintained or published separately from the index, standard thesaurus building practice has not been applied in establishing terms or references. For example, as noted in the introduction to the index, "'see' and 'see also' references are freely provided to lead the user to related entries." While LCSH is used as a major resource for identifying and selecting candidates for NUCMC subject terms, consistency with LCSH form or usage has not been a goal. In some cases, related LCSH terms are combined because fine distinctions in topics (e.g., between SLAVES and SLAVERY) cannot be applied to collections of manuscripts. In other cases, terms are combined for greater pre-coordination, since NUCMC is accessible only as a printed list. However, the three most significant variations from LCSH are driven by the nature of historical records and the research they support. The major variations may be characterized as follows:

1. Geographic access. Access by place name is essential to local history research. Since records about a place may be voluminous, NUCMC further subdivides place names by topic, thus

"doubling" the heading under topic that is further subdivided by place. If NUCMC topical headings were to be aligned with LCSH, it might still prove necessary to maintain a NUCMC place name index that included topical subdivisions.

2. Form headings. A precise term characterizing the form of historical records (e.g., Farm accounts) is necessary in describing archival collections. NUCMC currently uses a number of form headings and subdivisions not included in LCSH. In machine-readable cataloging of manuscript collections (a relatively new development), many of these terms could be used in MARC field 655. A separate controlled vocabulary (Form Terms For Archival and Manuscripts Control, MARC code "ftamc") is available as a source list for form terms. This list is currently being revised to make it more satisfactory for manuscript cataloging.

3. Terminology reflecting historical concepts. While LCSH must strive to eliminate outdated terminology, some outdated terms may provide more accurate descriptions of time-bound historical records. While "Teachers' colleges" is a term preferable (and used for) "Normal schools" in LCSH, the terms are not synonymous to individuals researching educational history. Thus, NCUMC often adheres to the principle of successive rather than latest form in its selection of terminology.

5.1.6 Handbook of Latin American Studies

The Handbook of Latin American Studies (HLAS) is an annual bibliography of books, journals, and journal articles on Latin America. The Handbook covers humanities and social sciences in alternate years, with each annual volume containing 4,000-5,000 citations. The citations are identified and prepared by some 120 contributing editors (most are working scholars) and are compiled and edited by the Hispanic Division of Research Services. One of the editorial tasks is the provision of a subject index for each year's volume. Each citation receives an average of 2-3 subject access points.

Less than 1 FTE is devoted to HLAS subject indexing. Since the index must be produced quickly after the citations are compiled in order to meet a publication deadline, little time is available for standardizing thesaurus maintenance practices or editorial review of new terms. The emphasis is on production of a publication; the HLAS thesaurus has no users outside of the HLAS editorial process. While a typed (word processed) thesaurus of HLAS terms and references was compiled a few years ago, it has not been fully maintained. The most authoritative list is considered to be the indexes from the two previous years, and indexing is done using the past publication as an authority file and guide as well as the thesaurus.

HLAS subject indexing employs a thesaurus that has been tailored to the needs of a published index and a defined area of scholarship. While LCSH is routinely checked as a source of new

terms and references, many non-LCSH terms are added to reflect the specificity of HLAS journal article coverage and its geographical approach. Periodical literature, which makes up 50% of each HLAS volume, often refers to new concepts and political groupings not yet reflected in the monographic literature covered by LCSH. Also, since the HLAS subject index does not cumulate, its universe of coverage each year is relatively small. Choice of terms and references is tailored to the coverage of each year's volume, rather than to the broad realm of citations covered by LCSH. "See also" references are made liberally to broader, narrower and related terms, and do not necessarily conform to LCSH or standard thesaurus practice. When an alternative form of an LCSH term is selected for HLAS use, a "see" reference is normally made from the LCSH form.

The differences between HLAS terms and LCSH are difficult to specify. However, they can be generally characterized as follows:

1. HLAS uses geographic subdivision for almost all terms since it is geographically oriented. Topical subjects are subdivided geographically and geographic areas are subdivided by topic.

2. Terminology will follow common usage in the contributors' fields, even when this requires deviation from LCSH.

3. LCSH may not contain a term used in Latin American studies, either because the term is too specific (e.g., "Minifundios") or the usage would be inconsistent with LCSH practice (e.g., "Mesoamerica")

4. HLAS uses direct word order rather than inverted terms.

5. Because its universe of coverage is much smaller than that of LCSH, HLAS often combines concepts that are split into separate terms in LCSH. (e.g., HLAS uses "Family and family relations" in lieu of a variety of separate terms used in LCSH.)

For the most part, the nature of differences between HLAS index terms and LCSH can be attributed to the development of the one list within the context of a particular area of scholarship and the other aimed to cover a broad universe of subjects. The lists are driven further apart by their distinctly different applications.

5.2 Computer Support for Thesaurus Management

LEXICO and LCSH Online, the two thesaurus maintenance systems currently used by the Library of Congress, are representative of the two kinds of thesaurus management system applications, that is, lexicographic systems and subject authority systems. The LEXICO system provides such features as automated maintenance of hierarchies, automatic creation of reciprocal references (i.e., cross-posting of references when the one-way relationship is provided), KWIC and KWOC displays of

terms. A number of different thesauri can be maintained on the system; an introductory menu directs users to the thesaurus file of choice. The different thesauri are maintained as separate databases; no cross-file links or combined products are supported. LEXICO is currently used to maintain LIV and TGM. It could also be used by the Handbook of Latin American Studies should the Handbook staff decide that maintenance of separate subject thesaurus would be a cost-effective or desirable investment. It is less likely that LEXICO would be well suited to maintaining the subject lists for NUCMC and the Annotated Card Program, since these lists are essentially exceptions to LCSH and not complete in themselves.

Some enhancements to LEXICO would be desirable to increase its effectiveness for its current users at LC. A significant current limitation is its minimal profiling for print output, since the system does not now produce a print tape adequate for the production of the published LIV thesaurus. A more critical limitation is the lack of a machine-readable output useable in other LC files. Thus LIV terms must be keyed twice: once for maintenance on LEXICO and again into the local batch system which produces both print tapes and a LIV file searchable on SCORPIO. In the long run, the lack of a MARC Authorities Format output or some other kind of interface to LC's bibliographic system would prevent thesauri created on LEXICO from being used as subject authority files.

LCSH Online, on the other hand, has been designed as a component of a larger bibliographic system; it comprises MUMS Subject Release 1.0. However, at present, the thesaurus maintenance features of LCSH Online are quite limited and MUMS authority files are not yet linked to bibliographic records. Since the initial version of the system varies little from LC Name Authorities implementation, the lexicographic needs of LCSH editors are not yet met. Enhancement aimed specifically at thesaurus editing would improve the quality of the LCSH as well as aid the work of catalogers and editors. For example, alphabetical KWIC displays would help editors maintain consistent terminology; hierarchical displays are crucial to evaluating new NT/BT relationships. The thesaurus management features described in Chapter Two may prove helpful in considering future enhancements for LCSH maintenance.

However, if LCSH Online were to evolve into LC's thesaurus maintenance/subject authority control system, some significant redesign would need to be considered if the system were to support thesauri in addition to LCSH. Effective thesaurus management requires the logical integrity of each list that is maintained. For example, any automatic creation and validation of reciprocal references must be restricted within each list. It is also necessary to view alphabetical and hierarchical listings for only one thesaurus at a time. In the current MUMS system, the subject source list (byte 11 of the MARC Authorities 008, FFD box 38 of the MUMS internal format) does not operate as a partition of the indexes. Segregated displays of different

thesauri require the maintenance of separate MUMS authority files for each. For smaller lists, this may not be a cost effective alternative.

Even if separate MUMS authority files were created (or loaded from LEXICO) for HLAS, LIV, and TGM, the "partial" lists maintained for NUCMC and ACP would continue to pose a problem. These lists are dependent upon LCSH; complete separate files would require duplication of LCSH terms used by NUCMC and ACP. Alternatively, creation of references, and eventually bibliographic headings validation, for NUCMC and ACP would require the use of two files--LCSH and the specialized file. (Note: This is a problem currently being addressed by UTLAS in its attempt to mount the Children's Headings file. Validation against a prioritized series of authority files--known as cascading in UTLAS terminology--is standard practice at UTLAS. However, UTLAS staff would prefer to maintain juvenile headings by mounting a complete Children's Headings file.) If LC wished to develop capabilities for handling multiple thesauri in a single subject authority file, a small addition to the MARC Authorities Format might aid in the support of the ACP and NUCMC lists. That is, an expansion of the 008 coding to show each of the subject source lists for which a term is valid. Such coding could--given appropriate applications design--obviate the need to duplicate LCSH records when the records are also valid for other derivative thesauri. However none of the authority control applications described in Chapter Three rely on such a device; all duplicate a term record when a term is used in more than one source list.

One last aspect of maintenance support needs to be considered in LC's multiple thesaurus environment. That is, the possibility of integrating separate thesauri by creating linking references between and among thesauri. Should LC determine that such links should be created, adequate thesaurus maintenance support for the links would be essential. Linked thesauri would need to be maintained on the same system (rather than, say, loading some from LEXICO to MUMS) and interfile displays and edit support as described in Chapter Two would be needed. For example, if some level of integration between LIV and LCSH were desirable, a specified related term relationship might be developed for terms postable in LCSH but a USE reference in LIV. A hit on a term so coded might result in a display constant such as, "Search also in BIBL under." A change on the LCSH term could result in a provisional change to the reference in LIV, with a report of the automatic change in the LIV reference made available for review by the LIV editor.

5.3 Online Subject Authority Control

The Library of Congress has not yet designed a system for linking its authority and bibliographic records. Implementation of online subject authority control will be a major undertaking, one that will affect the Library's cataloging system, its thesaurus and bibliographic file maintenance systems, and its

online catalog. The systems described in Chapter Three provide a menu of functions and file designs for LC's selection. The likely application of many of these features at LC depends in large part upon a key management question facing the Library's leadership. That is, to what extent is rethinking/redesign of MUMS and SCORPIO necessary, desirable, or possible in order to implement authority control?

If LC were to contemplate a significant redesign or replacement of existing bibliographic systems, authority control of multiple subject files need not pose a significant problem. The systems described in Chapter Three demonstrate that system designers have sucessfully met the challenge of multiple subject files. In fact, many of the functions and file structures necessary to support other desirable subject authority control features--e.g., subdivision validation, global change on subfields (not just unique strings), and links between authority records--are those which help to support the creation of multiple subject authority files. Given LC's complex, multi-file environment and the workload expended maintaining LCSH subdivisions, it would seem reasonable for the Library to look to sophisticated systems like Geac BPS and UTLAS as models for providing subject authority control.

If authority control implementation at LC proceeds within an environment of significant constraints on index design and file relationships, multiple subject authority files could pose problems. Since the Library creates and maintains subject authorities (i.e., thesauri) as well as subject headings, it would be important to maintain a logical separation of files. Subject indexes for different thesauri should be browsable both together and separately, either by partioning the indexes or by limiting searches to specific collections where the thesauri are used. In LC's applications, it would also be important to devise efficient means for copying or sharing subject authority records common to more than one thesaurus, since most NUCMC and ACP terms are in fact LCSH. Linking of headings to separate authority files should not pose a problem as long as the links are specified at the level of indicator and subfield and not just by MARC field tag. As noted in the preceding section, the minimum critical capability would be the availability and loading of LC's thesauri in a MARC/MUMS format.

5.4 Online Subject Searching

In contrast to achievements in authority control processing, no operational system has yet made the searching of terms from multiple thesauri in the same file "easy" for the untrained end-user. In determining how it will address this retrieval problem in its own catalogs, the Library of Congress will need to determine which of the four approaches outlined in Chapter Four it wishes to pursue: segregated files, mixed vocabularies, vocabulary integration, or front-end navigation.

Segregated files. The thesauri maintained by LC apply to a variety of specialized collections and databases, each in a different stage of automation. If the Library's online files are to begin to replace card catalogs for these collections, searches limited to those collections (and thus to those thesauri) will be necessary. While it is important that all of these databases be accessible to library users using the same logons and commands (i.e., in the same online catalog), how necessary is it to retrieve, for example, graphic materials, manuscripts, and books all in the same subject search? If special vocabularies have proven desirable for retrieving children's books or legislative information, is it also desirable to combine these materials in subject searching? If cross-file searching is not considered an essential option for general public access, the separate file approach may meet the Library's needs.

Mixed vocabularies. As noted in Chapter Four, the mixed vocabulary approach can be made more or less user-friendly depending on other significant design features of the subject access system. Since neither MUMS nor SCORPIO represents the state-of-the-art in online catalog subject searching, the question of possible system redesign becomes central to evaluating the mixed vocabulary approach for LC's catalogs. To the extent that redesign for subject authority control can affect indexes, syndetic structure, and displays, features can be added that could make mixed vocabulary searching more workable for the Library's users.

Chapter Four described some of the retrieval options that affect the clarity of results from a mixed vocabulary search. There is no obvious prescription for the Library of Congress. In the absence of successful operational mixed vocabulary catalogs or of relevant experimental research results, recommendations for mixed vocabulary searches must be based on opinion, common sense, and observations of the Library's users. Research has, however, demonstrated the value of subject index browsing to help users select terms. (7) It seems safe to predict that a well designed, readable subject browse that clearly identified the source of terms and references (probably in relation to collection rather than to thesaurus) would be a key requirement. This feature, coupled with well designed help screens and user training might make mixed vocabulary searching a viable means of access for LC's more sophisticated users (and/or its less discerning users as well). With the implementation of subject authority control, LC will also need to determine how to handle the non-LC subject terms (e.g., MeSH, NAL) it presently indexes in its online files.

Vocabulary integration. Because TGM, LIV, ACP, and NUCMC all use LCSH as a starting point for creating new terms, some degree of vocabulary integration could be feasible at the Library of Congress. Integrating the five thesauri would require additional computer-support features, including:

- common thesaurus management system and interfile searches and links as described in Chapter Two;

- subject authority control features that permitted specification and differentation of related term references across files (for example, a 550 relationship across authority files would need to be coded so that it did not conflict with an intrafile 450 reference from the same term);

- displays of retrieval results that clearly explained the relationship and sources of related terms from different thesauri.

However, no degree of computer support for integration would obviate the need for an increased intellectual effort to maintain the links between thesauri. While powerful thesaurus management software could free editorial time to be channelled into an integration effort, there are no data to indicate whether the investment would result in a greatly improved online catalog. The Library would need to decide whether this approach merits some experimentation.

Front-end navigation. Since this approach has not yet moved out of the realm of prototype systems, a proposed implementation of a "front-end interface" at the Library would be a longer range plan. It may also need to be based on redesign or replacement of MUMS/SCORPIO, rather than as an interface to the existing system. A workable strategy for LC would be to begin to explore the creation of such an interface while at the same time implementing a shorter-term solution, such as separately searched files. This strategy could have benefits beyond the Library of Congress, since the library profession (and ultimately library users) would be enriched by LC's entry into the field of online catalog research. Exercising the same leadership as it has in optical disk applications, LC could sponsor the development of a prototype "expert" system that would guide users in the selection of subject terms. The Library of Congress was one of the first to replace its century-old card catalogs with online access. The time may be right for LC to begin to develop the access system it will use in the 21st century.

Bibliography

1. Eisenberg, Michael. The Direct Use of Online Bibliographic Information Systems By Untrained End Users: A Review of Research. Syracuse, N.Y.: ERIC Clearinghouse on Information Resources, Syracuse University School of Education, 1983.

2. Hildreth, Charles R. Online Public Access Catalogs: The User Interface. Dublin, OH: OCLC, 1982.

3. Knapp, Sara D. "Creating BRS/TERM, a Vocabulary Database for Searchers," Database (December 1984) 70-75.

4. Kleinbart, Paul. "Prolegomenon to 'Intelligent' Thesaurus Software." Journal of Information Science (month 1985) 45-53.

5. Lancaster, F. Wilfrid and Smith, Linda C. Compatibility Issues Affecting Information Systems and Services. Paris: UNESCO General Information Program, 1983. (PGI-33/WS/23) 50-77.

6. LEXICO Thesaurus Management System: Users Manual Prepared for the Library of Congress. Bethesda, MD: Project Management, Inc., 1984.

7. Markey, Karen. Subject Searching in Library Catalogs Before and After the Introduction of Online Catalogs. Dublin, OH: OCLC, 1984.

8. MINISIS: Concepts and Facilities for Database Managers. Ottawa: International Development Research Centre, 1982.

9. Niehoff, Robert T. "Development of an Integrated Energy Vocabulary and the Possibilities for On-line Subject Switching." Journal of the American Society for Information Science 27 (Jan-Feb 1976) 3-17.

10. Niehoff, Robert and Mack, Greg. Evaluation of the Vocabulary Switching System: Final Report. Columbus, OH: Battelle Memorial Institute, 1984.

11. Niehoff, R.T. and Swasny, S. "The Role of Automated Subject Switching in a Distributed Information Network." Online Review 3:2 (1979) 181-194.

12. Peterson, Toni. Multiple Authorities in Library Systems. Unpublished paper. Presented at Art Libraries Society/North America Symposium on Authority Control, February 1986.

13. Sager, J.C., et al. "Thesaurus Integration in the Social Sciences." International Classification 8:3 (1981) 133-138; 9:1 (1982) 19-26; 9:2 (1982) 64-70.

14. Stam, Deidre. Choosing Our Words: Reflections on Authority Control in Art Information Systems. Unpublished paper, Presented at Art Libraries Society/North America Symposium on Authority Control, February 1986.

15. Taylor, Arlene G., et al. "Network and Vendor Authority Control." Library Resources and Technical Services 29 (April/June 1985) 195-205.

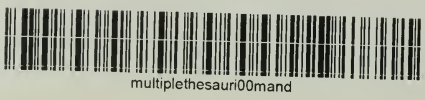
16. Wessells, Michael B. and Niehoff, Robert. "Synonym Switching and Authority Control." In Authority Control: The Key to Tomorrow's Catalog. Phoenix, AZ: Oryx, 1982, p. 97-118.

17. Williams, Martha E. "Electronic Databases." Science 228 (April 1985) 445-456.

BOSTON PUBLIC LIBRARY

multiplethesauri00mand

multiplethesauri00mand



multiplethesauri00mand

WITHDRAWN
No longer property of the
Boston Public Library.
Sale of this material benefits the Library